5962
13.7.82

# Psychological Bulletin

165

JAMES DEESE, Editor
*Johns Hopkins University*

Editorial Consultants

ROBERT ABELSON
JACK B. ADAMS
MAX M. ALLEN
DOUGLAS G. ANGER
MOSHE ANISFELD
DAVID L. ARENBERG
NATHAN AZRIN
WILLIAM P. BANKS
E. EARL BAUGHMAN
ALLEN E. BERGIN
DANIEL E. BERLYNE
WILLIAM BEVAN
SYDNEY W. BIJOU
ARTHUR J. BIOPELLE
HERBERT G. BIRCH
VERTUS E. BIXENSTINE
ROGER WADE BLACK
ELLIOTT BLASS
ROBERT C. BOLLES
KENNETH BONNET
ARTHUR H. BRAYFIELD
TIMOTHY C. BROCK
JOSEF BROZEK
M. P. BRYDEN
JAMES M. BUTCHER
DONALD T. CAMPBELL
VIRGIL R. CARLSON
PETER L. CARLTON
C. KEITH CONNORS
LEE J. CRONBACH
EDWARD E. CURETON
L. DOUGLAS DENIKE
CLINTON B. DESOTO
RALPH M. DREGER
HOWARD E. EGETH
DAVID EHRENFREUND
DAVID ELKIND
CHARLES W. ERIKSEN
BETSY WORTH ESTES
MARVIN J. FELDMAN
ROY E. FELDMAN
HAROLD J. FINE
THOMAS J. FISCHBACH
DONALD W. FISKE
GARLIE A. FOREHAND
FRANK FRIEDLANDER
SAMUEL C. FULKERSON
JOHN GAITO
GORDON GALLUP
RILEY W. GARDNER
SOL L. GARFIELD
ROBERT GLASER
GENE V. GLASS
GOLDINE C. GLESER
WALTER C. GOGEL
LEWIS R. GOLDBERG
SANDFORD GOLDSTONE
DONALD R. GOODENOUGH
BERT F. GREEN, JR.
SEBASTIAN P. GROSSMAN
A. RALPH HAKSTIAN
CURTIS D. HARDYCK
JOHN HEINZ
ECKHARD HESS
LARRY HOCHHAUS
PAUL J. HOFFMAN
ROBERT T. HOGAN
EDWIN P. HOLLANDER
EVELYN HOOKER
THOMAS J. HUMMEL
WILLIAM A. HUNT
STANLEY D. IMBER
PAUL D. ISAAC
ROBERT L. ISAACSON
MARVIN IVERSON
DOUGLAS N. JACKSON
ARTHUR B. JENSEN
GEORGE W. JOE
STEPHEN M. JOHNSON
MARSHALL B. JONES
SYDNEY M. JOURARD
ROBERT L. KAHN
JOE KAMIYA
STEPHEN KAPLAN
RICHARD E. KEESEY
DANIEL PORTER KIMBLE
H. D. KIMMEL
DAVID H. KRANTZ
JOHN D. KRUMBOLTZ
JULIUS LAFFAL
EDWARD LAUTER
HERSCHEL W. LEIBOWITZ
PETER M. LEWINSOHN
ALVIN M. LIBERMAN
EDWARD LICHTENSTEIN
LESTER LUBORSKY
DAVID L. MARGULES
G. ALAN MARLETT
ALBERT R. MARSTON
HAROLD G. MCCURDY
NORMAN ALVIN MILGRAM
JAMES G. MINARD
WALTER MISCHEL
HOWARD MOLTZ
JOHN MONEY
RUDOLF H. MOOS
PAUL MOUNTJOY
KENNETH E. MOYER
JAMES S. MYER
M. FRANK NORMAN
DAVID P. NOWLIS
VINCENT NOWLIS
JUM C. NUNNALLY
PAUL OBRIST
K. DANIEL O'LEARY
DAVID OLTON
GERALD R. PATTERSON
GORDON L. PAUL
DEAN PEABODY
H. S. PENNYPACKER
BENJAMIN POPE
LEO POSTMAN
HAROLD M. PROSHANSKY
RALPH M. REITAN
ROBERT A. RESCORLA
KLAUS F. RIEGEL
ERNEST Z. ROTHKOPF
HERBERT RUBINSTEIN
HALMUTH H. SCHAEFER
DUANE P. SCHULTZ
JEROME S. SCHWARTZBAUM
IRWIN SILVERMAN
HERBERT W. SIMON
EDWARD E. SMITH
KARL U. SMITH
CHARLES C. SPIKER
JULIAN C. STANLEY
CLIFFORD H. SWENSEN
KENNETH W. TERHUNE
DELBERT D. THIESSEN
WARREN TORGERSON
LEONA TYLER
THOM VERHAVE
ALLAN R. WAGNER
HERBERT J. WALBERG
ROBERT S. WALDROP
EDWARD L. WALKER
ROBERT WALKER
MICHAEL A. WALLACH
W. BRUCE WALSH
NANCY C. WAUGH
ROGER WEBB
KARL E. WEICK
WAYNE WICKELGREN
RUTH C. WILEY
WALLACE WILKINS
HERMAN A. WITKIN
MATTHEW YARCZOWER
JOSEPH L. ZINNES

## VOLUME 78, 1972

J962
13.7.82

HAROLD P. VAN COTT *Managing Editor*
JEANNE MCMANUS *Technical Editor*
HENRY S. SALMON *Advertising and Business Manager*
ANNE REDMAN *Subscription Manager*

PUBLISHED MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 SEVENTEENTH ST. N. W.
WASHINGTON, D. C. 20036

$10.25

Copyright © 1972 by the American Psychological Association, Inc.

# CONTENTS OF VOLUME 78

Althoff, Michael E. See Anthony, William A.

Anderson, Norman H. Looking for Configurality in Clinical Judgment ................ 93

Anderson, Norman H. See Weiss, David J.

Anthony, William A., Buell, Gregory J., Sharratt, Sara, and Althoff, Michael E. Efficacy of Psychiatric Rehabilitation ........................................ 447

Appelle, Stuart. Perception and Discrimination as a Function of Stimulus Orientation: The "Oblique Effect" in Man and Animals ........................................ 266

Atkinson, R. C., and Paulson, J. A. An Appoach to the Psychology of Instruction .... 49

Baddeley, A. D. Retrieval Rules and Semantic Coding in Short-Term Memory ........ 379

Beatty, William W. How Blind Is Blind? A Simple Procedure for Estimating Observer Naiveté ........................................ 70

Block, Jack. The Shifting Definitions of Acquiescence ........................................ 10

Bonnet, Kenneth. See Gaito, John.

Bovet, Daniel, Bovet-Nitti, Filomena, and Oliverio, Alberto. Gene Action and Behavior of C3H and DBA Mice: A Reply to Wilcock and Thiessen .................... 351

Bovet-Nitti, Filomena. See Bovet, Daniel.

Buell, Gregory J. See Anthony, William A.

Church, Russell M., and Getty, David J. Some Consequences of the Reaction to an Aversive Event ........................................ 21

Clark, Russell D., III, and Willems, Edwin P. Two Interpretations of Brown's Hypothesis for the Risky Shift ........................................ 62

Cureton, Edward E. See D'Agosino, Ralph B.

D'Agostino, Ralph B., and Cureton, Edward E. Test of Normality against Skewed Alternatives ........................................ 262

Davison, Gerald C., and Wilson, G. Terence. Critique of "Desensitization: Social and Cognitive Factors Underlying the Effectiveness of Wolpe's Procedure" ........... 28

Doherty, Michael E., and Keeley, Stuart M. On the Identification of Repeatedly Presented, Brief Visual Stimuli ........................................ 142

Eysenck, H. J. A Note on "Factors Influencing the Outcome of Psychotherapy" ...... 403

Fowler, Barry. Some Comments on "A Behavioral Approach to Nitrogen Narcosis" .. 234

Fried, P. A. Septum and Behavior: A Review ........................................ 292

Gaito, John, and Bonnet, Kenneth. Controls in Molecular Psychobiological Research .. 483

Getty, David J. See Church, Russell M.

Giambra, Leonard M. Comment on "Equivalence of Information in Concept Identification" ........................................ 186

Greenough, William T., and Maier, Steven F. Molecular Changes during Learning: Behavioral Strategy—A Comment on Gaito and Bonnet ........................... 480

Gynther, Malcolm D. White Norms and Black MMPIs: A Prescription for Discrimination? ........................................ 386

Heneman, Herbert G., III, and Schwab, Donald P. Evaluation of Research on Expectancy Theory Predictions of Employee Performance ........................................ 1

Hersen, Michel. Nightmare Behavior: A Review ........................................ 37

Hudson, Robert L. The Hudson-Dunn Clustering Index Revisited .................... 475

Kaplan, Bonnie J. Malnutrition and Mental Deficiency ........................................ 321

Keeley, Stuart M. See Doherty, Michael E.

Koulack, David. Rapid Eye Movements and Visual Imagery during Sleep ............. 155

Lawlis, G. Frank, and Lu, Elba. Judgment of Counseling Process: Reliability, Agreement, and Error ........................................ 17

Lenzer, Irmingard I. Differences between Behavior Reinforced by Electrical Stimulation of the Brain and Conventionally Reinforced Behavior ........................................ 103

LESTER, DAVID. Self-Mutilating Behavior .......... 119
LEVIN, JOEL R., AND MARASCUILO, LEONARD A. Type IV Errors and Interactions .......... 368
LEVY, JOHN. Autokinetic Illusion: A Systematic Review of Theories, Measures, and Independent Variables .......... 457
LOOFT, WILLIAM R. Egocentrism and Social Interaction across the Life Span .......... 73
LU, ELBA. See LAWLIS, G. FRANK.
LUBORSKY, LESTER. Another Reply to Eysenck .......... 406
MAIER, STEVEN F. See GREENOUGH, WILLIAM T.
MARASCUILO, LEONARD A. See LEVIN, JOEL R.
MARKLEY, O. W. Rejoinder to Sherif .......... 479
MARLATT, G. ALAN. Task Structure and the Experimental Modification of Verbal Behavior 335
MITCHELL, TERENCE R. See POLLARD, WILLIAM E.
MOLFESE, DENNIS L. See PALERMO, DAVID S.
MULHOLLAND, THOMAS B. Occipital Alpha Revisited .......... 176
O'CONNOR, EDWARD F., JR. Response to Cronbach and Furby's "How We Should Measure Change—Or Should We?" .......... 159
OLIVERIO, ALBERTO. See BOVET, DANIEL.
OWEN, DAVID R. The 47,XYY Male: A Review .......... 209
PALERMO, DAVID S., AND MOLFESE, DENNIS L. Language Acquisition from Age Five Onward .......... 409
PAULSON, J. A. See ATKINSON, R. C.
PERFETTI, CHARLES A. Psychosemantics: Some Cognitive Aspects of Structural Meaning 241
PETERS, EDWARD N. Nonexistent Individual Differences in Reminiscence .......... 375
POLLARD, WILLIAM E., AND MITCHELL, TERENCE R. Decision Theory Analysis of Social Power .......... 433
RICHARDSON, JOHN T. E. Nonparametric Indexes of Sensitivity and Response Bias .......... 429
RIEGEL, KLAUS F. Influence of Economic and Political Ideologies on the Development of Developmental Psychology .......... 129
SAMELSON, FRANZ. Response Style: A Psychologist's Fallacy? .......... 13
SCHOOLER, CARMI. Birth Order Effects: Not Here, Not Now! .......... 161
SCHWAB, DONALD P. See HENEMAN, HERBERT G., III.
SHARRATT, SARA. See ANTHONY, WILLIAM A.
SHERIF, CAROLYN W. Comment on Interpretation of Latitude of Rejection as an "Artifact" 476
SPITZ, HERMAN H. Note on Immediate Memory for Digits: Invariance over the Years .. 183
THEODOR, L. H. A Neglected Parameter: Some Comments on "A Table for the Calculation of d′ and β" .......... 260
VEROFF, JOANNE B. See VEROFF, JOSEPH.
VEROFF, JOSEPH, AND VEROFF, JOANNE B. Reconsideration of a Measure of Power Motivation .......... 279
WARREN, NEIL. African Infant Precocity .......... 353
WEINSTOCK, ROY B. Maintenance Schedules and Hunger Drive: An Examination of the Rat Literature .......... 311
WEISS, DAVID J., AND ANDERSON, NORMAN H. Use of Rank Order Data in Functional Measurement .......... 64
WEISS, STANLEY J. Stimulus Compounding in Free-Operant and Classical Conditioning: A Review and Analysis .......... 189
WILKINS, WALLACE. Desensitization: Getting It Together with Davison and Wilson .... 32
WILLEMS, EDWIN P. See CLARK, RUSSELL D., III.
WILSON, G. TERENCE. See DAVISON, GERALD C.

Date ....... 42 .........
d. N : ... Lib .........
Bureau Edul Psy. Research.

VOL. 78, NO. 1 JULY 1972

Psychological Bulletin

# EVALUATION OF RESEARCH ON EXPECTANCY THEORY PREDICTIONS OF EMPLOYEE PERFORMANCE

HERBERT G. HENEMAN III[1] AND DONALD P. SCHWAB[2]

*Graduate School of Business and Industrial Relations Research Institute, University of Wisconsin*

Formulations of expectancy theory proposed by L. W. Porter and E. E. Lawler in 1968 and by V. H. Vroom in 1964 are summarized. Nine field studies testing expectancy theory hypotheses are then reviewed. Finally, the studies are evaluated in terms of the number of independent variables used, the measurement of these variables, and the statistical analysis performed. It is shown that the research has generally been inadequate in all three respects. Suggestions for improvement in expectancy theory research are made throughout the review.

A central concern of industrial relations is the identification and measurement of factors associated with individual differences in employee job performance, since efficient utilization of manpower resources is dependent upon our ability to account for such differences. Recently, two statements of a theory of employee job performance, usually referred to as expectancy theory, have been advanced by Vroom (1964) and Porter and Lawler (1968). These formulations have stimulated considerable thought and research. Unfortunately, it appears that there is substantial confusion in interpretation of, and conduct of research on, the theory. Our purposes are thus to attempt to resolve this state of confusion and make suggestions for improvement in future research on expectancy theory.

In subsequent sections, Vroom's (1964) and Porter and Lawler's (1968) descriptions of the theory are summarized and shown to be quite similar. Nine field studies that have tested hypotheses of expectancy theory are the summarized. Finally, these studies are critically evaluated and recommendations are made for needed additional research.

## THE THEORY

Vroom (1964) hypothesized that employee job performance (P) is a function of the interaction between force to perform (F), or motivation, and ability (A).

$$P = f(F \times A). \qquad [1]$$

Porter and Lawler (1968), alternatively, hypothesized that performance is a function of the three-way interaction among exerted effort (E), or motivation, ability, and role perceptions (R).

$$P = f(E \times A \times R). \qquad [2]$$

Conceptualizations of motivation, ability, and role perceptions are described below. While terminological differences exist between the two theories, the only substantive difference between them is the inclusion of role perceptions as an additional performance determinant by Porter and Lawler. For this reason, and for purposes of clarification, the theories are combined into one comprehensive framework, shown in Figure 1.

### *Force or Effort*

In Vroom's formulation, force to exert effort level $i$ is a function of the sum of the in-

---

[1] Requests for reprints should be sent to Herbert G. Heneman III, Graduate School of Business, 1155 Observatory Drive, University of Wisconsin, Madison, Wisconsin 53706.

[2] The authors thank R. J. Adams, L. L. Cummings, and W. E. Scott, Jr., for their critical comments on an earlier draft of this paper.

© 1972 by the American Psychological Association, Inc.

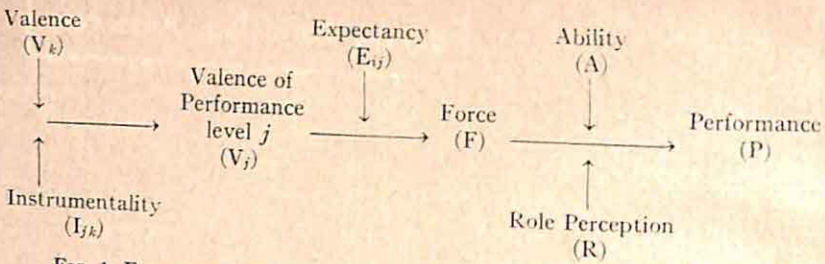


FIG. 1. Expectancy theory predictions of job performance, based upon Vroom (1964) and Porter and Lawler (1968).

teractions between (*a*) the valence of each performance level ($V_j$) and (*b*) the perceived probability that an *i*th amount of effort will result in the achievement of each performance level; this perception is termed expectancy ($E_{ij}$).

$$F_i = f[\sum_{j=1}^{n} (V_j \times E_{ij})]. \qquad [3]$$

It is hypothesized that the individual chooses the effort level exerting the strongest positive or weakest negative force.

A performance level (first-level outcome), in turn, acquires valence if it is perceived as leading to the attainment of valent second-level outcomes ($V_k$), such as pay and recognition. The perception that performance level *j* will result in the attainment of second-level outcome *k* is termed the instrumentality of the performance level ($I_{jk}$). The valence of performance level *j* is hypothesized to be a function of the sum of the interactions between (*a*) the valence of second-level outcomes and (*b*) the instrumentalities of the performance level.

$$V_j = f[\sum_{k=1}^{n} (V_k \times I_{jk})]. \qquad [4]$$

Porter and Lawler (1968) presented a theoretical treatment of motivation to perform, further refined by Lawler (1970), which is similar to Vroom's except for differences in terminology. Vroom's terminology, and the corresponding terminology by Porter and Lawler, is as follows: force (effort), valence of second-level outcomes (value of rewards), instrumentality (performance-reward probability), and expectancy (effort-performance probability). The latter two form what Porter and Lawler term the effort-reward probability.[3] For expository convenience the terminology of Vroom is employed throughout the rest of this discussion.

### *Ability*

Both Vroom (1964) and Porter and Lawler (1968) hypothesized that the individual's actual performance is additionally dependent upon his ability to perform. Vroom defined ability as characteristics of the individual that represent "a potential for performing some task which may or may not be utilized [p. 198]." Similarly, Porter and Lawler (1968) defined ability as "relatively stable, long-term individual characteristics (e.g., personality traits, intelligence, manual skills, etc.) that represent the individual's currently developed power to perform [p. 22]." In both models emphasis is placed on the individual's capacity, as opposed to his willingness, to perform at a task.

### *Role Perceptions*

Porter and Lawler (1968) additionally included role perceptions as a determinant of performance. Role perceptions are defined as

---

[3] Another apparent discrepancy between Vroom (1964) and Porter and Lawler (1968) is that the latter first multiply the effort-performance and performance-reward probabilities, and then multiply this product and value of rewards to form their definition of effort. In the case of only one second-level outcome, this procedure results in a definition of effort identical to Vroom's (1964) definition of force. With more than one second-level outcome, however, this procedure leads to a definitional discrepancy between Vroom and Porter and Lawler. Consistency with Vroom requires that the value of reward and performance-reward probabilities be multiplied and summed across second-level outcomes, and that this total product be multiplied by the effort-performance probability. In his latest statement of expectancy theory, Lawler (1970) did combine the variables in this manner, and thus is consistent with Vroom.

"the direction of effort—the kinds of activities and behaviors the individual believes he should engage in to perform his job successfully [p. 24]." Evaluation of these behaviors by his supervisor, however, is dependent upon the supervisor's role perceptions for the job. This suggests that accuracy of role perceptions may be a more important performance determinant than the individual's role perceptions per se.

## THE EVIDENCE

Nine published field studies were found that investigated one or more hypotheses of expectancy theory using various measures of employee performance as the dependent variable. In reviewing these studies the terminology previously developed is employed. Reference to significant results indicates statistical significance at or beyond the .05 level. Critical evaluation of the studies is reserved for the following section since a number of the comments apply to more than one study.

Georgopolous, Mahoney, and Jones (1957) preceded Vroom's (1964) and Porter and Lawler's (1968) formulations by testing a portion of expectancy theory on 621 manufacturing operatives. Productivity was hypothesized to be a function of path-goal perceptions (i.e., instrumentality), level of need (i.e., valence), and freedom to alter one's performance behavior. Instrumentality and valence perceptions were obtained for three second-level outcomes. Freedom consisted of self-reported freedom, age, and length of service. Productivity was measured by subjects' reports of their productivity relative to established standards. A contingency problem format (Blalock, 1960) was employed to compute main and "interaction" effects of the three independent variables. Generally, significant differences and "interactions" were obtained for the outcome "making more money in the long run." For example, the percentage of high producers was significantly greater for the high money need–positive path-goal perception group than the high money need–negative path-goal perception group.

Lawler (1966) investigated the effects of instrumentality and ability on the performance of 211 state government managers. Subjects were asked to indicate the instrumentality[4] of three factors (quality, productivity, and effort) for the determination of their pay. Responses to these items were summed to form a composite instrumentality index. Ability was measured by supervisor rankings of managerial qualifications. Supervisor rankings and self-ratings of performance served as dependent variables. The independent variables were dichotomized and a two-way analysis of variance was performed for each dependent variable. The main instrumentality effect was significant for both dependent variables; ability and interaction effects were each significant for one dependent variable.

Galbraith and Cummings (1967) obtained ratings of the valence of seven second-level outcomes, and the instrumentality of performance for their attainment, from 32 operatives in a manufacturing organization. Ability was measured by length of time on the job, and performance was measured by daily output as a percentage of standard averaged over a 1-month interval. The independent variables were dichotomized, coded in dummy variable format (including interaction variables), and subjected to stepwise multiple regression analysis. No valence, instrumentality, or ability variables entered any of the regression equations significantly, and only 6 of 54 possible interaction variables entered the equations significantly.

Lawler and Porter (1967) examined the effects of valence, instrumentality, role perceptions, and their interactions on the performance of 154 managers. Subjects were asked to indicate how instrumental three factors (effort, high productivity, and good job performance) were for the attainment of seven second-level outcomes and the valence of these outcomes. Responses to the instrumentality items were summed to form a composite instrumentality index. Role perceptions were obtained from 55 of the managers by having them rank 10 inner- and outer-directed traits (Riesman, 1950; Whyte, 1956) first on the basis of self-description and second on the basis of importance for success. Six criteria (three ratings of effort and three of performance) were obtained.

---

[4] In this and several other studies (Gavin, 1970; Hackman & Porter, 1968; Lawler, 1968a; Lawler & Porter, 1967; Porter & Lawler, 1968), individuals were asked to indicate the relationship between "working hard" or "effort" and second-level outcomes. Their responses presumably represent the perceived effort-reward probability. As such, they confound the effort-performance (expectancy) and performance-reward (instrumentality) probabilities. For purposes of simplicity, however, such perceptions are referred to as instrumentality perceptions in this review.

The median correlation between instrumentality and (*a*) performance was .11 and (*b*) effort was .18. Correlations between valence and the criteria were not reported. Out of 42 possible correlations between valence times instrumentality scores and the criteria, 29 were larger than the correlations between instrumentality and the criteria alone, 5 were the same, and 8 were smaller. Finally, high and low valence times instrumentality groups were formed, and correlations between role perceptions and performance were computed. Some of the correlations within groups were significant, but the correlations between groups were not significantly different. The results thus offer some support for the effects of role perceptions, but no support for the interaction of valence instrumentality and role perceptions.

Hackman and Porter (1968) obtained ratings of the instrumentality of working hard for the attainment of 18 second-level outcomes, and the valence of each outcome, from 82 female service representatives of a telephone company. Four summated scores were computed for each subject: valence, instrumentality, valence plus instrumentality, and valence times instrumentality. The scores were correlated with 10 subcriteria of job performance and a composite performance criterion. The median correlation between the 10 subcriteria and (*a*) valence was .16; (*b*) instrumentality was .11; (*c*) valence plus instrumentality was .17; (*d*) valence times instrumentality was .27. The correlation between valence times instrumentality and the composite criterion was a significant .40. Since the valence-instrumentality scores yielded greater correlations with performance than the other sets of scores, Hackman and Porter argued that the results tend to confirm the hypothesized valence-instrumentality interaction effect.

Lawler (1968a) used cross-lagged (Campbell & Stanley, 1963) and dynamic (Vroom, 1966) correlation analyses to test for causality between valence-instrumentality perceptions and performance. Data were collected from 55 public service managers twice, with a 1-year interval between collections. Subjects indicated how instrumental two factors (working hard and quality of job performance) were for attaining six second-level outcomes and the valence of each outcome. Six composite valence-instrumentality scores were computed for each subject. Three measures of job performance were obtained. The cross-lagged correlation analysis suggested that valence instrumentality was more likely the cause, rather than the result, of job performance. The dynamic correlation analysis, however, generally did not preclude the possibility that some third variable was responsible for the results obtained from the cross-lagged analysis.

Porter and Lawler (1968) investigated the effects of valence of pay, instrumentality, and role perceptions on the performance of 635 managers. These variables were measured as in the Lawler and Porter (1967) study. Dependent measures were self-ratings and supervisory rankings of effort and performance. Hypotheses were tested by the significance of difference between mean performance and effort of managers giving the highest and lowest third of responses to the independent variable(s) in question. In general, the results offered some support for the main effects of instrumentality and role perceptions. Less support was obtained for hypothesized interaction effects.

Gavin (1970) investigated the impact of valence, instrumentality, ability, and role perceptions on the performance of 192 male and 175 female managerial candidates of an insurance company. Each subject was asked to indicate the instrumentality of two factors (good job performance and working hard) for attaining 21 second-level outcomes, and the valence of each outcome. A total composite valence-instrumentality score was computed for each subject. Based upon the work of Barrett (1966), two measures of role perceptions were obtained by the correlations between (*a*) the subject's and his supervisor's rating of the importance of 22 work behaviors for the subject's performance and (*b*) the subject's importance and self-descriptiveness ratings of the same behaviors. Ability was measured by a general "in-house" mental aptitude test. Performance was measured by supervisory ratings. Of the independent variables, only ability did not correlate significantly with performance. Multiple regression procedures suggested by Cohen (1968) also were used. Additive regression models generally yielded significant multiple correlations on the order of .30. Interaction variables generally did not significantly "contribute to" these multiple correlations, thus indicating little support for hypothesized interaction effects.

Goodman, Rose, and Furcon (1970) interviewed 66 employees of a government research organization to obtain indications of the three most valent second-level outcomes, the instrumentality of seven factors for attaining these outcomes, and the amount of job control (in order to infer expectancy). A measure of force (see Formula 3) was derived for each of the seven factors. Actual measures of success on four of these seven factors were obtained and correlated with the appropriate force measure. Significant correlations were obtained in each instance.

### *Summary*

Nine field studies that tested hypotheses derived from expectancy theory were reviewed. Generally, valence, instrumentality, and role perceptions were significantly related to performance, while ability was not. Little support was obtained for hypothesized interactions among these variables.

## DISCUSSION OF THE EVIDENCE

The review suggests that limitations of existing research pertain primarily to the variables investigated and their measurement and to the analytical procedures employed. Measurement criticisms are aimed at each major independent[5] variable in the theory (force, ability, and role perceptions). In the analysis section, research objectives are specified, and the analyses performed in several of the studies are criticized in light of these objectives.

### *Force to Perform*

Expectancy theory clearly places greatest emphasis on the role of motivation or force to perform as a determinant of job performance. It is disconcerting to observe, therefore, that force has consistently been measured incorrectly in the studies reviewed. Either instrumentality (performance-reward probability) and expectancy (effort-performance probability) has not been clearly delineated, or only instrumentality has been measured directly. For example, Hackman and Porter (1968) confounded instrumentality and expectancy perceptions by asking respondents to specify the probable linkage between effort and second-level outcomes. Gavin (1970), Lawler (1966, 1968a), Lawler and Porter (1967), and Porter and Lawler (1968) obtained overall perceived effort-reward and performance-reward probabilities. In each study the two probabilities were combined, again resulting in confounding. In the remaining three studies (Galbraith & Cummings, 1967; Georgopolous et al., 1957; Goodman et al., 1970) only instrumentality perceptions were measured directly. Attempts were made to infer expectancy by considering the nature of the job on such dimensions as autonomy and self-control, but subjects were never asked to indicate directly their expectancy perceptions.

Additionally, in five of the studies (Gavin, 1970; Lawler, 1966, 1968a; Lawler & Porter, 1967; Porter & Lawler, 1968) instrumentality alone has been confounded. According to the theory, the perceived instrumentality of each first-level outcome should be related separately to measured success on that outcome (e.g., as done by Goodman et al., 1970). For example, if "quality of job performance" is the first-level outcome, instrumentality perceptions of "quality of job performance" should be related empirically to actual "quality of job performance." Such a procedure was not followed in the five studies identified. Rather, managers rated the instrumentality of more than one first-level outcome, the ratings were combined to form a composite instrumentality index, and the composite index was related to actual success on various first-level outcomes.

---

[5] Discussion of the dependent measure, performance, is omitted because the problem permeates a variety of research problems in organizational settings and because excellent treatments are presented by Campbell, Dunnette, Lawler, and Weick (1970), Dunnette (1966), Guion (1965), Nagle (1953), and Ronan and Prien (1966).

Turning to valence, the number and type of second-level outcomes used in most of the studies may be questioned. Research on job satisfaction (e.g., Herzberg, Mausner, & Snyderman, 1959; Smith, Kendall, & Hulin, 1969; Vroom, 1964; Weiss, Dawis, England, & Lofquist, 1967) suggests that individuals experience a wide variety of satisfactions in work environments. These may be sorted (either conceptually or empirically) into a number of categories or factors, each of which may be viewed as a second-level outcome derived from participation in the organization and/or performance in a work role. A meaningful measure of force thus should tap valence and instrumentality perceptions for all relevant outcomes. The relevance of outcomes to the individuals under investigation, however, is difficult to determine a priori. This suggests that many outcomes should be included in the measure developed, or that preliminary investigation be conducted to determine the relevant second-level outcomes for the group studied. Unfortunately, the trend has been to include few second-level outcomes without preliminary investigation.

Finally, the problem of stability of responses appears to be serious, although only two studies have investigated this type of reliability. Galbraith and Cummings (1967) obtained a test–retest reliability of .80 for instrumentality, but only .50 for valence over a 1-month interval. Lawler (1968a) reported a test–retest reliability of .48 for valence-instrumentality responses over a 1-year interval.

### *Ability*

In only three studies has ability even been considered as an explanatory variable. Lawler (1966) measured ability by having supervisors rank subordinates on overall qualifications. This ranking correlated significantly with the supervisor's ranking of his subordinates on overall job performance. Criterion contamination in some degree is likely, however, since both rankings were obtained from the same supervisor at the same time. Galbraith and Cummings (1967) defined ability as length of time on the job. The extent to which length of time on the job serves as a proxy for ability as defined, however, is at best unclear. Gavin (1970) used a psychometric ability measure. It did not correlate significantly with performance, nor did its interactions with force and role perceptions generally contribute significantly to the multiple correlations. In part, this was probably due to restriction of range since the measure was employed as a selection instrument by the organization. Gavin also argued that the measure may not tap the relevant intellectual capacities.

Future research requires experimentation with numerous psychometric ability measures. Use of these measures should be guided by the validity evidence of various aptitude and achievement tests used for predicting employee performance in a selection context (e.g., Ghiselli, 1966; Guion, 1965). In addition, given the broad definition of ability presented by Vroom (1964) and Porter and Lawler (1968), measures of interest, temperament, and personality also might be considered. Their use requires caution, however, for they generally have not correlated significantly with performance (Dunnette, 1966; Guion & Gottier, 1965; Nash, 1965), and they may tap motivational characteristics of individuals (Guion, 1965).

### *Role Perceptions*

The effects of role perceptions have also been investigated in only three instances. Lawler and Porter (1967) and Porter and Lawler (1968) asked managers to rank a number of psychological traits in the order of perceived importance for determining success in their present position. By defining the traits with one word it is doubtful that even approximately equal meanings were triggered for all respondents (Stryker, 1958). Moreover, no attempt was made to tap role perception accuracy as defined by Porter and Lawler (1968).

Gavin's (1970) role perception measures appear to represent an improvement over Porter and Lawler's since Gavin employed more behaviorally descriptive items and included accuracy. Nevertheless, because the items that formed the basis of the measures were not developed with specific reference to the managerial positions investigated by Gavin, their relevance and completeness of description are open to question.

More attention should be devoted to accuracy of role perceptions in future research. Additionally, it seems that more meaningful measures would concentrate on the inclusion of the job duties and objectives derived from detailed analysis of the specific job under consideration. It is these duties and requirements that ultimately define the individual's job and form the basis of his performance evaluation.

### *Statistical Analysis*

Expectancy theory hypothesizes that performance is a function of several independent variables operating in an interactive fashion. Thus, it is reasonable to assume that analytical procedures performed in tests of the theory would indicate (*a*) the amount of performance variance explained by additive and interactive variables and (*b*) the extent to which hypothesized interactions among variables made independent, significant contributions to explained performance variance beyond their additive (main) effects (i.e., the "usefulness" of interaction variables, Darlington, 1968).

The procedure employed by Gavin (1970) is the best to date in terms of these analytical objectives. He regressed performance on the additive variables, and then the additive and interactive variables, to obtain estimates of performance variance explained and the usefulness of the interactions investigated. This procedure thus has much to commend it in future research on expectancy theory.

Lawler (1966) also obtained interpretable indications of interaction effects by performing a two-way analysis of variance. Dichotomization of independent variables, however, may have led to an understatement of effects. Moreover, computation of omega squared (Hays, 1963) would have been desirable to indicate the strength of main and interaction effects.

Lawler and Porter (1967), and Hackman and Porter (1968), correlated valence, instrumentality, or both with performance. Valence and instrumentality scores were then multiplied, and the product score was correlated with performance. Generally, the latter correlations exceeded the former. The significance of the increase was not reported, however, and the usefulness of the valence-instrumentality interaction is thus indeterminant. Goodman et al. (1970) and Lawler (1968a) merely correlated valence-instrumentality scores with performance, and thus there is no indication of the usefulness of those interactions.

Galbraith and Cummings (1967) stepwise regressed performance against all independent variables, which were dichotomized. Two problems may be noted with this approach. First, the reduction of continuous variables to dichotomous ones may have lowered possible predictability. Second, such a procedure does not necessarily indicate the usefulness of interaction variables. If an interaction variable enters a stepwise regression before one or more of its component variables, it is impossible to demonstrate that the interaction variable made a unique contribution to the multiple correlation beyond the additive effects of its component variables.

In the remaining studies (Georgopolous et al., 1957; Porter & Lawler, 1968) significance of difference tests were performed on groups created by dichotomizing independent variable scores. Interactions, in turn, were assessed by performing significance of difference tests on the dependent variables between levels of one independent variable, within levels of one or more other independent variables. Such techniques are suboptimal for several reasons. First, dichotomization may result in an underestimate of possible effects. Second, no direct indications about the strengths of relationships are given. Third, it is difficult to precisely determine and convey interaction effects. Finally, a substantial number of significance tests are required, resulting in possible nonindependence of significance tests and an underestimate of the number of Type I errors.

## CONCLUSIONS

Expectancy theory differs in several promising respects from much previous theorizing about the determinants of employee performance. Campbell et al. (1970), for example, pointed out that the concept of force to perform focuses on the major classes of variables (valence, instrumentality, and expectancy) that determine motivated behavior. This is in direct contrast to other industrial motivation theories (e.g., Herzberg et al., 1959; Maslow, 1965) that tend to emphasize only identification of second-level outcomes. Expectancy theory additionally includes ability and role perceptions as potential performance determinants. This multivariate approach seemingly has potential for increasing the proportion of explained variance in employee performance. Finally, interactions among independent variables are explicitly hypothesized, and presumably these interactions will account for significantly greater performance variance than additive formulations alone.

Unfortunately, these potential theoretical improvements have not been adequately reflected in the research. First, in only one study was a measure of force to perform developed according to its theoretical definition. Second, in only one study were measures of all three independent variables (force, ability, and role perceptions) included. The predictive power of the total theory is thus essentially unknown. Finally, while various interactions were supposedly computed in all of the studies, in only two was the contribution of the interactions to explained performance variance, beyond additive effects, appropriately computed. A minimum requirement for future research is thus that it avoid replication of these three general inadequacies in past research.

At the level of specific measurement issues, the review and discussion of the evidence warrants the following conclusions and suggestions for future research. First, in terms of force to perform, efforts should be made to incorporate an adequate number and type of second-level outcomes, and to develop measures of both instrumentality and expectancy that are more reflective of their corresponding constructs as defined by Vroom (1964) and Porter and Lawler (1968). Second, psychometrically sound ability measures should be employed. Finally, the development and use of behaviorally descriptive role-perception measures, and assessment of the impact of role perception accuracy, are required.

A noteworthy aspect of research on expectancy theory is the emphasis on investigating employees in their natural work environments, thus providing a high degree of external validity. In the case of motivation, this is in direct contrast to research on equity theory (Adams, 1965) and goal-setting theory (Locke, 1968), which has usually entailed student subjects working on laboratory tasks in experimental situations. The "cost" of external validity has been, of course, a general inability to make causal inferences. Lawler's (1968a) design exemplifies one potential compromise between the needs to study employees in their natural work environments and to obtain evidence that permits causal inferences. A variety of such designs are suggested by Campbell and Stanley (1963), and hopefully they will be used more fully in future research.

While there is an obvious need for substantial additional research on the validity of expectancy theory as formulated, work should also be performed on the determinants of valence, instrumentality, and expectancy perceptions. It has been hypothesized, for example, that the valence of second-level outcomes is a function of the satisfaction derived from their past attainment (Porter & Lawler, 1968) and their perceived equity (Lawler, 1968b). A number of specific variables have been hypothesized to influence instrumentality. In general, they all pertain to the nature of the actual relationships between performance and second-level outcomes, on the assumption that these "objective" relationships will influence instrumentality perceptions. This approach has been most noticeable in the discussions of compensation systems (Opsahl & Dunnette, 1966; Schneider & Olson, 1970). Recently it has been extended to considerations of job design as a means of more directly relating performance and intrinsic second-level outcomes (Lawler, 1969); and to considerations of the supervisor as a motivational agent to the extent that he can differentially reward the performance of his subordinates (Evans, 1970). In terms of expectancy determinants, Lawler (1970) hypothesized that past expectancy experiences and self-esteem influence expectancy. In addition, the nature of the task on such dimensions as difficulty, control over work pace, and length of job cycle may be important determinants. Considerably more theoretical and empirical work of this nature is necessary for the development of an expanded model of force to perform.

## REFERENCES

ADAMS, J. S. Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965.

BARRETT, R. S. The influence of supervisor's requirements on ratings. *Personnel Psychology,* 1966, **19**, 375–387.

BLALOCK, H. M. *Social statistics*. New York: McGraw-Hill, 1960.

CAMPBELL, D. T., & STANLEY, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally, 1963.

CAMPBELL, J. P., DUNNETTE, M. D., LAWLER, E. E., & WEICK, K. E. *Managerial behavior, performance, and effectiveness*. New York: McGraw-Hill, 1970.

COHEN, J. Multiple regression as a general data-analytic system. *Psychological Bulletin,* 1968, **70**, 426–443.

DARLINGTON, R. B. Multiple regression in psychological research and practice. *Psychological Bulletin,* 1968, **69**, 161–182.

DUNNETTE, M. D. *Personnel selection and placement.* Belmont, Calif.: Wadsworth, 1966.

EVANS, M. G. The effects of supervisory behavior on the path-goal relationship. *Organizational Behavior and Human Performance,* 1970, **5**, 277–298.

GALBRAITH, J., & CUMMINGS, L. L. An empirical investigation of the motivational determinants of task performance: Interactive effects between valence-instrumentality and motivation-ability. *Organizational Behavior and Human Performance,* 1967, **2**, 237–257.

GAVIN, J. F. Ability, effort, and role perception as antecedents of job performance. (Experimental publication system, manuscript number 190A) Washington, D.C.: APA, 1970.

GEORGOPOLOUS, B. S., MAHONEY, G. M., & JONES, N. W. A path-goal approach to productivity. *Journal of Applied Psychology,* 1957, **41**, 345–353.

GHISELLI, E. E. *The validity of occupational aptitude tests*. New York: Wiley, 1966.

GOODMAN, P. S., ROSE, J. H., & FURCON, J. E. Comparison of motivational antecedents of the work performance of scientists and engineers. *Journal of Applied Psychology,* 1970, **54**, 491–495.

GUION, R. M. *Personnel testing*. New York: McGraw-Hill, 1965.

GUION, R. M., & GOTTIER, R. F. Validity of personality measures in personnel selection. *Personnel Psychology,* 1965, **18**, 135–164.

HACKMAN, J. R., & PORTER, L. W. Expectancy theory predictions of work effectiveness. *Organizational Behavior and Human Performance,* 1968, **3**, 417–426.

HAYS, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. *The motivation to work*. New York: Wiley, 1959.

LAWLER, E. E. Ability as a moderator of the relationship between job attitudes and job performance. *Personnel Psychology,* 1966, **19**, 153–164.

LAWLER, E. E. A correlational—causal analysis of the relationship between expectancy attitudes and job performance. *Journal of Applied Psychology,* 1968, **52**, 462–468. (a)

LAWLER, E. E. Equity theory as a predictor of productivity and work quality. *Psychological Bulletin,* 1968, **70**, 596–610. (b)

LAWLER, E. E. Job design and employee motivation. *Personnel Psychology,* 1969, **22**, 426–435.

LAWLER, E. E. Job attitudes and employee motivation: Theory, research, and practice. *Personnel Psychology,* 1970, **23**, 223–238.

LAWLER, E. E., & PORTER, L. W. Antecedent attitudes of effective managerial job performance. *Organizational Behavior and Human Performance,* 1967, **2**, 122–142.

LOCKE, E. A. Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance,* 1968, **3**, 157–189.

MASLOW, A. *Eupsychian management*. Homewood, Ill.: Irwin-Dorsey, 1965.

NAGLE, B. F. Criterion development. *Personnel Psychology,* 1953, **6**, 271–289.

NASH, A. N. Vocational interests of effective managers: A review of the literature. *Personnel Psychology,* 1965, **18**, 21–37.

OPSAHL, R. L., & DUNNETTE, M. D. The role of financial compensation in industrial motivation. *Psychological Bulletin,* 1966, **66**, 94–118.

PORTER, L. W., & LAWLER, E. E. *Managerial attitudes and performance*. Homewood, Ill.: Irwin-Dorsey, 1968.

RIESMAN, D. *The lonely crowd*. New Haven, Conn.: Yale University Press, 1950.

RONAN, W. W., & PRIEN, E. P. *Toward a criterion theory*. Greensboro, N.C.: Richardson Foundation, Creativity Research Institute, 1966.

SCHNEIDER, B., & OLSON, L. K. Effort as a correlate of organizational reward system and individual values. *Personnel Psychology,* 1970, **23**, 313–326.

SMITH, P. C., KENDALL, L. M., & HULIN, C. L. *The measurement of satisfaction in work and retirement*. Chicago: Rand-McNally, 1969.

STRYKER, P. On the meaning of executive qualities. *Fortune,* 1958, **57**, 116–119.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

VROOM, V. H. A comparison of static and dynamic correlation methods in the study of organizations. *Organizational Behavior and Human Performance,* 1966, **1**, 55–70.

WEISS, D. J., DAWIS, R. V., ENGLAND, G. W., & LOFQUIST, L. H. *Manual for the Minnesota Satisfaction Questionnaire*. Minneapolis: Industrial Relations Center, University of Minnesota, 1967.

WHYTE, W. H. *The organization man*. New York: Simon and Schuster, 1956.

(Received October 8, 1970)

Psychological Bulletin
1972, Vol. 78, No. 1, 10–12

# THE SHIFTING DEFINITIONS OF ACQUIESCENCE[1]

JACK BLOCK[2]

*University of California, Berkeley*

The Bentler, Jackson, and Messick rebuttal is not judged to be responsive to the criticisms leveled earlier. The shifting over the years in the definition of acquiescence and in the kinds of evidence used to bolster the acquiescence interpretation is noted. In particular, it is noted that highly specific and replicable evidence is required to support the presently advanced hypothesis that acceptance acquiescence importantly influences assessment procedures. The Bentler et al. view that *any* evidence of response inconsistency is evidence for their "distinct species of acquiescence" is not accepted.

Readers of *Psychological Bulletin* are doubtless satiated by the obdurate insistence of the principals on debating the usefulness of the acquiescence concept. I continue to hold up my side of the argument because I believe the field of personality assessment will advance more wisely in the future if it is allowed to gain clarity on its recent past. In my present remarks on a response (Bentler, Jackson, & Messick, 1972) to a critique (Block, 1971) of a conjecture (Bentler, Jackson, & Messick, 1971) that was revised after heavy criticism (e.g., Block, 1965; Lichtenstein & Bryan, 1965; Rorer, 1965; Rorer & Goldberg, 1965; Samelson, 1964; Samelson & Yates, 1967) of the initial proposal (e.g., Jackson & Messick, 1958; Messick & Jackson, 1961), I shall be terse and declaratory. The ultimate judgment on this prolonged controversy will come, not from further discussion, but from the way subsequent research is shaped.

The counterings by Bentler et al. (1972) to my earlier critique (Block, 1971) do not, I believe, require much further comment. Only the Morf and Jackson (1972) study has been proposed as positive evidence for the revised conception of acquiescence. The studies by Bentler (1969) and by Bentler and Marshall[3] are proposed by Bentler et al. as evidence only for "noncontent responding," and the reader is then asked to interpret "noncontent responding" as evidence for acquiescence.

The positive evidence, the Morf and Jackson (1972) study, even after correction and reanalysis, continues to depend on items rejected earlier by Jackson because of their mediocre content validity; on brief a priori scales that are impressively unreliable and entirely unvalidated; on the arbitrary injection of large amounts of unanchored variance into a factor analysis by using unities as the communality estimates for these chancy scales when an average communality of only .28 is justified; and on a Procrustean rotation of scale vectors close to the point of origin of the factor space and therefore highly conformable to any factor target matrix. I do not believe our understanding of why an individual responds to an inventory statement as he does is especially advanced by the Morf and Jackson (1972) study. For me, this sequence of analyses and assumptions is simply too removed, too arcane, and too aleatoric[4] to be helpful.

---

[1] Preparation of this manuscript was facilitated by United States Public Health Service Research Grant MH-16080.

[2] Requests for reprints should be sent to Jack Block, Department of Psychology, University of California, Berkeley, California 94720.

[3] P. M. Bentler and M. Marshall. Effect of presentation format on item validity. Unpublished manuscript, University of California, Los Angeles, 1969.

[4] The newly proposed method for evaluating factor score reliability (Jackson & Morf, 1971), cited by Bentler et al. (1972) as supportive of the significance of the Morf and Jackson (1972) factors, is logically faulty and does not address itself to the question, Will the Morf and Jackson factors fit a randomly constructed factor target matrix as well as the factors fit the hypothesized target matrix? It seems to me that Morf's [M. E. Morf. An analysis of two response styles: True responding and item endorsement. Unpublished doctoral dissertation, University of Western Ontario, 1969] inability to find, in his original data, a factor representing acceptance acquiescence when he partitioned his subject samples into male and female subsamples, despite efforts to

The larger purpose of the present note, however, is not to discourse upon the particulars of the Bentler et al. (1972) rebuttal but rather to comment on changes that have occurred over the years in the way the advocates of the importance of acquiescence have construed the concept and sought to evidence it. By so doing, the inconsistency that Bentler et al. (1971, 1972) view me as manifesting can be reconciled, or at least placed in a perspective that I would prefer.

In the heyday of acquiescence, some 10 or 12 years ago, there was only one kind of acquiescence which, in the current Bentler et al. (1971, 1972) terminology, is identifiable as agreement acquiescence. Agreement acquiescence then was viewed as "play[ing] a dominant role in personality inventories like the MMPI [Messick & Jackson, 1961, p. 303]." It was further hoped that agreement acquiescence would prove to be a basic variable underlying individual differences, a variable that could integrate a diversity of behaviors.

Where are we now? It is acknowledged by acquiescence adherents (Bentler et al., 1971, p. 191) that agreement acquiescence no longer can be viewed as important for understanding the MMPI. Empirical efforts to establish agreement acquiescence as a broad and encompassing dimension underlying individual differences have failed. The field of personality assessment has turned to other concerns. The new inventory, the Personality Research Form (Jackson, 1967a), hardly mentions acquiescence and has elected to obviate the entire issue by the tactic of balancing scales (after Block, 1965).[5]

But despite this history, or perhaps because of it, acquiescence adherents have shifted their claims for its importance to another variety, acceptance acquiescence. Acceptance acquiescence was first invoked by Jackson and Messick (1965) in a brief report as a conjecture that, if true, might explain a Rorer and Goldberg (1965) study that had severely embarrassed the concept of agreement acquiescence. A more elaborate and formal version of the acceptance acquiescence notion was presented by Bentler et al (1971). Time will tell whether the notion of acceptance acquiescence, conceived in response to the established insufficiency of agreement acquiescence and evidenced to date only by the shaky Morf and Jackson (1972) factor analysis, can find further and substantial support. I am content to wait.

Meanwhile, however, it should be noted that the highly specific conjecture regarding acceptance acquiescence offered by Bentler et al. (1971) requires empirical confirmation of a highly specific and configured nature. Demonstrations that the validity of inventory and assessment measures is less than perfect, that is, indications of "noncontent responding," simply do not provide the special evidence that is required if we are to accept the existence of acceptance acquiescence. Assessment measures can be, and are, imperfect for a wide variety of reasons—inadequate scale development, lexically ambiguous or contrary items, time constraints imposed on respondents, idiosyncratic self-evolved sets, vague or undesirable instructions, item format deficiencies, naive psychometric models, various categorizing proclivities, pure and irreducible error, and so on, all confounded together.

For those of us interested in *using* personality inventories and other assessment procedures, no particular purpose is served by employing items, scales, or measures that fail to try to reduce "noncontent responding." Just as the experimenter wishes to prevent irrelevant sources of variance from arising in his experiment and can do so by the application of general principles of experimental design, the personality assessor wishes to exclude, at the outset, unnecessary or uninteresting or possibly confounding response manifestations so that his assessment procedure

---

maximize fit to this hypothesis and despite the strong part–whole relationships between his subsamples and his total sample, testifies to the ready frangibility of the factor fit reported for the sample when the sexes are combined. It will be recalled that it is acceptance acquiescence that is the important kind of acquiescence now.

[5] It merits attention that although Jackson (1967b) as well as Messick [S. Messick. Psychology and methodology of response styles. Paper presented at the annual meeting of the Western Psychological Association, Honolulu, Hawaii, June 1965] have argued strongly that balancing scales is a poor way of excluding such acquiescence as conceivably might be present, Jackson's inventory employs balanced scales.

will be valid for its purpose and also concise. He can do so by applying a variety of psychometric and psychological principles—rigorous scale development, ambiguity elimination, appropriate item formats, development of unequivocal and desirable instructions and subject orientations, sophisticated psychometric models, forced choices, and so on. My advocacy of methodological controls over the last 15 years or so (cf., e.g., Block, 1956), which is seen by Bentler et al. (1972) as inconsistent with my slighting of the importance of acquiescence, can thus be seen to be simply and congruently motivated.

In their closing paragraph, Bentler et al. stated, "What we have referred to as distinct species of acquiescence, Block prefers to call response styles [Bentler et al., 1972, p. 112]." This statement and its elaboration suggest to me that Bentler et al. are unable or unwilling to recognize the existence of response tendencies that have no connection whatsoever with their "distinct species of acquiescence." It follows that, as long as they insist upon equating all "noncontent responding" with their "species of acquiescence," Bentler et al. will be formally correct in interpreting the fallibility of assessment devices as evidence for acquiescence. Indeed, given this definition, acceptance and agreement acquiescences in complex but unspecified interaction can be said to be everywhere. Those of us who reject this all-embracing and magisterial interpretation will not be impressed by these claims but will continue to seek a more differentiated and a more psychological understanding of the responses of humans in assessment situations.

## REFERENCES

Bentler, P. M. Semantic space is (approximately) bipolar. *Journal of Psychology*, 1969, **71**, 33–40.

Bentler, P. M., Jackson, D. N., & Messick, S. The identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, 1971, **76**, 186–204.

Bentler, P. M., Jackson, D. N., & Messick, S. A rose by any other name. *Psychological Bulletin*, 1972, **77**, 109–113.

Block, J. A. A comparison of the forced and unforced Q-sorting procedures. *Educational and Psychological Measurement*, 1956, **16**, 481–493.

Block, J. *The challenge of response sets.* New York: Appleton-Century-Crofts, 1965.

Block, J. On further conjectures regarding acquiescence. *Psychological Bulletin*, 1971, **76**, 205–210.

Jackson, D. N. *Personality Research Form Manual.* Goshen, N. Y.: Research Psychologists Press, 1967. (a)

Jackson, D. N. Review of J. Block, The challenge of response sets. *Educational and Psychological Measurement*, 1967, **27**, 207–219. (b)

Jackson, D. N., & Messick, S. Content and style in personality assessment. *Psychological Bulletin*, 1958, **55**, 243–252.

Jackson, D. N., & Messick, S. Acquiescence: The nonvanishing variance component. *American Psychologist*, 1965, **20**, 498. (Abstract)

Jackson, D. N., & Morf, M. An empirical evaluation of factor reliability. (University of Western Ontario Research Bulletin No. 199) London: Ontario: 1971. (On file, Archives Division, University of Western Ontario Library)

Lichtenstein, E., & Bryan, S. H. Acquiescence and the MMPI: An item reversal approach. *Journal of Abnormal Psychology*, 1965, **70**, 290–294.

Messick, S. J., & Jackson, D. N. Acquiescence and the factorial interpretation of the MMPI. *Psychological Bulletin*, 1961, **58**, 299–304.

Morf, M. E., & Jackson, D. N. An analysis of two response styles: True responding and item endorsement. *Educational and Psychological Measurement*, 1972, in press.

Rorer, L. G. The great response-style myth. *Psychological Bulletin*, 1965, **63**, 129–156.

Rorer, L. G., & Goldberg, L. R. Acquiescence in the MMPI? *Educational and Psychological Measurement*, 1965, **25**, 801–817.

Samelson, F. Agreement set and anticontent attitudes in the F scale: A reinterpretation. *Journal of Abnormal and Social Psychology*, 1964, **68**, 338–342.

Samelson, F., & Yates, J. F. Acquiescence and the F scale: Old assumptions and new data. *Psychological Bulletin*, 1967, **68**, 91–103.

(Early publication received January 28, 1972)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 13–16

# RESPONSE STYLE:

## A PSYCHOLOGIST'S FALLACY?

FRANZ SAMELSON[1]

*Kansas State University*

The recent attempt by Bentler, Jackson, and Messick to revive the acquiescence hypothesis is analyzed. It is shown that (*a*) they fail to clarify the conceptual meaning of response style; (*b*) their operational definition of content responding and acquiescent styles employs, as did earlier studies of acquiescence, an incorrect logical model—incorrect since it treats all opposites as formally equal instead of distinguishing between contraries and contradictories. While consequently (*c*) an acquiescence interpretation of their data is gratuitous, (*d*) these data fail to meet Bentler et al.'s own criteria for response style. The root of the problem seems to be that the experimenters mistake their own interpretation, based on an oversimplified model, for the (actual or functional) item content and then define all discrepant responses as acquiescence—a mistake of the sort called "the psychologist's fallacy" by William James.

Recently, an article by Bentler, Jackson, and Messick (1971) attempted to resurrect the "acquiescence" hypothesis. Block (1971) has already replied to some of their points. But since Bentler et al. (1971) view the basic issue as one of data, it might be useful to show that they persist in an inadequate analysis of the problem, that is, of the conceptual meaning of response style and the operational definition of acquiescence.

### WHEN ARE RESPONSES A STYLE?

The initial, straightforward conception of response bias or style as content independent ran into empirical difficulties, producing successively more ephemeral redefinitions. Bentler et al. (1971) now "attempt to clarify the role of stylistic variables and of specific item content . . . [p. 187—note the dichotomy]" in determining responses. This clarification takes three forms: some examples, an abstract definition, and a series of "conclusions." On inspection, the first two appear to be contradictory, and the third useless.

Quoting Jackson and Messick (1962), Bentler et al. presented the abstract definition of response style as stating three criteria: (*a*) "relatively enduring over time . . . ," (*b*) "generality beyond a particular test performance . . . ," and (*c*) "consistencies in response to item characteristics other than specific content [p. 188]." The *examples* of response style describe three subjects saying "true" to inventory items, for reasons that seem to be either situation specific (job interview, group pressure), or to invoke a personality construct (undifferentiated self-concept) of questionable relevance and at best adumbrating the general idea. In short, the examples may confuse rather than clarify.

### WHEN IS A TRAIT A STYLE?

The nine "conclusions" are not very helpful either. The first one, defining style as a personality variable or trait (p. 188), gives us the genus of the concept. But we look in vain for the differentia of the class of styles. The other conclusions list various style attributes that seem common to fairly broad personality characteristics, for example, authoritarianism and intelligence. But little seems gained were we to call such phenomena response styles; some differences might be lost, as by confusing authoritarianism and acquiescence.

Only the abstract definition provides any differentiae, by referring to expressive consistencies, found in response to item characteristics other than specific content. But this outcome of our search leaves us exactly where we started, with the dichotomy (between stylistic variables and specific content) to be clarified. The crucial terms remain undefined; we have to guess at their meaning. And it seems that, apart from the vague notion of expressiveness, the authors' distinction reflects an experimenter's bewilderment upon confronting the unforeseen emergence of global response patterns from an item pool, which he expected would measure some discrete and face-valid traits. Indeed, the Bentler et al. article seems to say so explicitly in referring to patterns with "a basis, not in the specific personality variable *of interest to the scale constructor* [italics added], but in other expressive consistencies . . . [p. 188]"; and implicitly, since the revival of response style seems to originate in the unexpected correlations of Bentler's (1969) semantic differential study. Is this the clue to the meaning of response style?

---

[1] Space limitations required adoption of a more abrupt and dogmatic style of presentation than may be desirable. Copies of the original paper containing some elaboration and documentation, as well as reprints of the present version, are available from F. Samelson, Department of Psychology, Kansas State University, Manhattan, Kansas 66502.

## ARE WE ALL EITHER SAINTS OR SINNERS?

Bentler et al. (1971) emerge from this conceptual impasse with an operational definition of two response styles: agreement and acceptance acquiescence. It consists of correlation patterns obtained from sets of items worded in opposite directions, together with a logical model generating patterns of content or style responses (p. 190).

But the essential feature of this model is not new; in fact, most agreement set studies with the F Scale were based on it. As shown earlier (Samelson & Yates, 1967), one of its main assumptions is wrong: that formally all opposites are equal—when in fact some are more opposite than others. By neglecting the old logical distinction between contraries and contradictories, and then forcing responses into a—for the task—inappropriate dichotomy (true–false, present–absent), some strange correlations can be produced. Failing to fit simplistic content-based expectations, they are then labeled "response style." This seems to be happening again in the present case—which is not surprising, since the problem goes unnoticed or at least unmentioned.

Since the basic idea is rather simple, let us illustrate it rather than spelling it out formally. Bentler et al. (1971) constructed their model in terms of the opposites *happy, sad, not happy, not sad*. They correctly expect me to reply "false" to *sad*, after I said "true" to *happy*. They are wrong in expecting, as they do, that I must say "true" to *sad* after saying "false" to *happy*. There is quite a distance between these (contrary) opposites. To assume that I am not responding to the "specific content" if I say "false" to both is turning things upside down. I have failed to respond only to what *they* see (without telling me) as the content: a pair of contradictories.

Similarly, meaningful answers to the pairs *happy–not sad*, etc., are possible, although they may contribute to negative correlations where the model postulates a positive one, and vice versa. (Only two of the six pairs, *sad–not sad* and *happy–not happy*, look like genuine contradictories. Yet even here the complexities of natural language and real people may not altogether fit into a formal model.)

What can happen to correlations between items or scales in such situations has been spelled out in an earlier article (Samelson & Yates, 1967). The main point is that one cannot expect the patterns of correlations postulated by Bentler et al. (1971) for content-based responses; nor can one infer agreement or acceptance style from any such pattern (unless it is extreme). The difference between style and content turns out again to be in the experimenter's expectations; misled by the abstract form given to the actual situation, he demands too simple a result from content responses, and calls it a mixture of response style and content when the data come out differently.

## "IT IS DIFFICULT TO FIND MANY SPECIFIC *SCALES* WHICH ARE ORTHOGONAL WITH RESPECT TO EVALUATION"[2]

Unexpected correlation patterns seem to lie at the bottom of the response style idea. And at first glance, the results of Bentler's (1969) study of semantic space may seem surprising. Besides the failure of "opposites" to correlate negatively, the study showed positive correlations between supposedly independent semantic dimensions. Given only the correlation matrix, one can but guess at their meaning. But Bentler (1969) provided some details in his article, among them sets of sample adjectives used to represent the "dimensions." For example, for the positive pole of Activity, he listed *active, alive, dynamic, moving, lively*. Is it really impossible to understand, on a content basis, why Evaluation correlates .65 with Activity represented by these adjectives—which read like a list of good old American values? The negative pole of Potency is represented by *smooth, delicate, easy, simple, light*. Are we really forced into an acquiescence interpretation after finding a positive correlation between such adjectives and Evaluation?

---

[2] From Osgood, Suci, and Tannenbaum (1957, p. 38).

Osgood et al. made numerous references to the difficulty of separating evaluation from other dimensions: "The noticeable tendency for both activity and power to be associated with positive evaluation . . . may represent a cultural semantic bias [p. 38]." Or "Evaluation thus appears as a highly generalizable attribute which may align itself with almost any other dimension of meaning . . . [p. 188]." And this problem is aggravated by Bentler's (1969) use of single adjectives instead of pairs that had allowed Osgood et al. some control for evaluative connotations.

## HOW TO PRODUCE RESPONSE SETS—NOT STYLES

Finally, a general point on method may be relevant. In Bentler's (1969) study, 200 subjects had to rate a concept on a list of 532 adjectives. For Morf and Jackson (as cited in Bentler et al., 1971), 195 subjects had to respond to a 560-item inventory plus an adjective checklist (in two sessions). In Bentler and Marshall's study (also cited in Bentler et al., 1971), 200 subjects responded "true–false" to 600 adjectives. It seems a reasonable guess that such conditions should be optimal for the adoptions of response sets by overtaxed subjects. ("It becomes curiously hard, after you have studied the applicability of two hundred adjectives to a given concept to think of what possible attributes might even apply to that concept [Deese, 1970, p. 109].")

As briefly reported by Bentler et al., Trott and Jackson found an increase in content-inconsistent responses when items were projected for short intervals (down to 3 seconds per item). A similar result appeared when Jackson and Lay reduced the "content saturation" of their items. While the base rate may have been improperly diagnosed as "acquiescence," the rate *increase* may well have represented failure to respond to item content. As we know from earlier studies, showing a blank screen is bound to produce responses independent of item content.

But such results, and all others in the Bentler et al. article—even if accepted as showing responses unrelated to specific item content—fail to meet their own criteria for a response *style:* consistencies enduring over time, generality beyond particular test performances, and "ordinarily . . . correlations with certain other behavioral consistencies . . . [p. 188]." None of the results reported involve enduringness, generality, and relations to other behaviors.

Thus the article fails to clarify its own question and seems to provide irrelevant answers. As before, "evidence" for acquiescence may not be in the data but may be imposed on them by an incorrect model. A more fine-grained analysis of what individuals do with specific items seems required before we have the right to attribute traits of illogic and/or acquiescence to our defenseless subjects. And new "style" arguments have to break out of the correlation matrix of "true–false, agree–disagree" responses to show their usefulness for our understanding of something more substantial. Until then, the acquiescence idea may deserve a good rest. It was an idea fruitful (as the saying goes) in generating research. But William James warned us long ago of "the great snare of the psychologist . . . [:] *the confusion of his own standpoint with that of the mental fact* about which he is making his report. . . . The most fictitious puzzles have been introduced into our science by this means [1890, p. 196]."

## REFERENCES

BENTLER, P. M. Semantic space is (approximately) bipolar. *Journal of Psychology*, 1969, 71, 33–40.

BENTLER, P. M., JACKSON, D. N., & MESSICK, S. Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin,* 1971, **76**, 186–204.

Block, J. On further conjectures regarding acquiescence. *Psychological Bulletin,* 1971, **76**, 205–210.

Deese, J. *Psycholinguistics.* Boston: Allyn and Bacon, 1970.

Jackson, D. N., & Messick, S. Response styles and the assessment of psychopathology. In S. Messick & J. Ross (Eds.), *Measurement in personality and cognition.* New York: Wiley, 1962.

James, W. *The principles of psychology.* New York: Holt, 1890.

Osgood, C. G., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

Samelson, F., & Yates, J. F. Acquiescence and the F Scale: Old assumptions and new data. *Psychological Bulletin,* 1967, **68**, 91–103.

(Early publication received March 2, 1972)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 17–20

# JUDGMENT OF COUNSELING PROCESS:
## RELIABILITY, AGREEMENT, AND ERROR

G. FRANK LAWLIS[1] AND ELBA LU

*Arkansas Rehabilitation Research and Training Center, University of Arkansas*

The purpose of this article is to present a solution in meeting the problem of defining and measuring agreement of judges and raters as used in psychological research. Because reliabilities and correlations are measures of consistency rather than agreement, a chi-square analysis is presented and probabilities are discussed to reveal the degree of agreement or disagreement. By using this method, the researcher can predetermine his own criterion and compare the frequencies of agreements with expected probabilities.

Rating scales are being used with increasing frequency to study relevant therapeutic interpersonal skills from tape-recorded therapy sessions. The use of such rating scales carries with it certain methodological considerations. Whenever a group of judges is presented a stimulus to which they must respond with a numeric interval that designates a level of functioning, an issue revolves around whether or not the judges can reach a unanimous perception.

In the past, the agreement of judges has been often equated to the reliability of ratings and analyzed accordingly. The purpose of this article is to demonstrate that agreement and reliability of ratings are different concepts and need different measuring techniques, and to propose a coefficient of agreement that can be adopted under general conditions and adapted to the experimenter's needs.

The general approach in judging reliabilities of ratings has been to account for proportional consistency of variance to raters. This consistency has been reported either in analysis of variance or correlational statistics. These approaches may yield quite high reliability coefficients although the agreement is clearly questionable, as can be seen from the following data (see Tables 1 and 2) that are examined by means of Ebel's (1951) coefficient and the product-moment correlation coefficient (Ferguson, 1959). In both cases, the estimated reliability is quite high and yet, by informally viewing the data, obviously only in one instance do the raters agree completely.

Another prime disadvantage of both approaches to the reliability of ratings is their lack of power to detect relationships when the total variance of range is small, as is often the case for psychotherapy scales. In Table 3, the series of ratings have a small range of two, and, although it appears that judges are fairly close in agreement, the product-moment correlation coefficient is relatively low (−.15). When two scores are added that drastically increase the range, the correlation is significantly increased (.91).

The proposed technique is basically a nonparametric method of computing agreement and disagreement according to a criterion predetermined by the investigator. The following notation is needed. Let $N$ be the number of items or individuals being rated, $N_1$ the number of observed agreements, $N_2$ the number of observed disagreements, $k$ the number of judges, $n$ the number of points on the scale, and $p$ the probability of $k$ judges achieving agreement on an individual by chance.

The model for computing the probability $p$ is based upon a rectangular distribution, that is, every judgment has the same probability of occurring under the hypothesis that the judges have no understanding of the scale being applied and their ratings are purely random. When this is true, $p$ is equal to the number of possible ways of obtaining agree-

---

[1] Requests for reprints should be sent to G. Frank Lawlis, who is now at Department of Psychology, Texas Technological University, P. O. Box 4100, Lubbock, Texas 79409.

TABLE 1

RATINGS OF 10 PERSONS OBTAINED FROM TWO RATERS FOR A CERTAIN TRAIT, PREPARED FOR COMPUTATION OF EBEL'S RELIABILITY COEFFICIENT Σ

| Person | Rater |  | $\Sigma X_p$ | $(\Sigma X_p)^2$ |
|---|---|---|---|---|
|  | A | B |  |  |
| 1 | 4 | 3 | 7 | 49 |
| 2 | 5 | 3 | 8 | 64 |
| 3 | 5 | 4 | 9 | 81 |
| 4 | 1 | 1 | 2 | 4 |
| 5 | 2 | 1 | 3 | 9 |
| 6 | 3 | 1 | 4 | 16 |
| 7 | 4 | 2 | 6 | 36 |
| 8 | 6 | 5 | 11 | 121 |
| 9 | 7 | 5 | 12 | 144 |
| 10 | 8 | 6 | 14 | 196 |
|  | $\Sigma X_r = 45$<br>$(\Sigma X_r)^2 = 2025$<br>$\Sigma X^2 = 372$ | $\Sigma X_r = 31$<br>$(\Sigma X_r)^2 = 961$ | $\Sigma X = 76$<br>$\Sigma(\Sigma X_r)^2 = 2986$ | $\Sigma(\Sigma X_p)^2 = 720$ |

ment divided by $n^k$, the number of possible combinations of $k$ ratings. The null hypothesis of agreement by chance can now be tested by using the statistics

$$\chi^2 = \frac{(N_1 - Np - .5)^2}{Np} + \frac{(N_2 - N(1-p) - .5)^2}{N(1-p)}$$

where .5 is a correction for continuity. The sampling distribution of this chi-square statistic is a chi-square distribution with one degree of freedom.

The test proceeds now in the usual manner. At a specified significance level, if the null hypothesis cannot be rejected, the conclusion is that there is no indication of nonrandomness in the ratings, and it would be advisable to question further the applicability of the scale. If the null hypothesis is rejected, the conclusion is that the judges do have discriminating power, and the observed agreement (or disagreement) is not due to chance.

TABLE 2

SUMMARY TABLE

| Source | *df* | *SS* | *MS* |
|---|---|---|---|
| Raters | 1 | 9.80 | — |
| Persons | 9 | 71.20 | 7.91 |
| Error | 9 | 2.20 | .24 |
| Total | 19 | 83.20 |  |

Depending upon the rigorousness of a particular project, one of the following criteria is usually chosen as the basis for determining agreement on an item being judged on a psychotherapy scale: (*a*) Let $r = 0$, identical rating; (*b*) let $r = 1$, deviation of 1 point; (*c*) let $r = 2$, deviation of 2 points.

For the first criterion ($r = 0$), the probability of judges reaching agreement by chance

TABLE 3

RATINGS ON A CERTAIN TRAIT BY TWO JUDGES (FICTIONAL DATA) JUDGED BY THE PRODUCT-MOMENT CORRELATION COEFFICIENT

| Person | Rater |  | Product-moment correlation coefficient |
|---|---|---|---|
|  | A | B |  |
| 1 | 4 | 3 |  |
| 2 | 2 | 3 |  |
| 3 | 4 | 2 |  |
| 4 | 4 | 3 |  |
| 5 | 4 | 4 |  |
| 6 | 3 | 4 |  |
| 7 | 3 | 3 |  |
| 8 | 4 | 3 |  |
| 9 | 9 | 9 | −.15 |
| 10[a] | 1 | 1 | .91 |

[a] Addition of two scores increases range.

alone is given by the following equation: $p = (1/n)^{k-1}$. For the second criterion ($r = 1$) the equation would be:

$$p = \frac{(n-1)\sum_{i=1}^{k-1} 2^{k-1} + n}{n^k}.$$

The third criterion is chiefly used when there are many points on a scale and one wishes the criterion of agreement to be broader. In that case ($r = 2$), the formula for $p$ is:

$$p = \frac{3^{k-2}(5n-6) + S\sum_{i=3}^{k} 3^{k-i}(n-3)2^{i-1}}{n^k}$$

where $S$ is a characteristic function such that $S = 1$ if $k \geq 3$; $S = 0$, if $k < 3$.

Table 4 represents the probability $p$ of $k$ judges agreeing by chance on each of the three different criteria. This is computed for a scale of 5 points and for a scale of 10 points.

TABLE 4

PROBABILITIES OF CHANCE AGREEMENT BASED ON RECTANGULAR MODEL

| Judges | Criterion | | |
|---|---|---|---|
| | $r = 0$ | $r = 1$ | $r = 2$ |
| | 5-point scale | | |
| $k = 2$ | .200 | .5200 | .7600 |
| $k = 3$ | .040 | .2320 | .5200 |
| $k = 4$ | .008 | .0976 | .3376 |
| | 10-point scale | | |
| $k = 2$ | .100 | .2800 | .4400 |
| $k = 3$ | .010 | .0640 | .1600 |
| $k = 4$ | .001 | .00136 | .0536 |

In Table 5, 50 segments of psychotherapy sessions recorded on tape were analyzed on the dimension of therapeutic warmth. The scale applied has five levels of warmth that fit the assumption of a rectangle distribution. Using a criterion of agreement of $r = 1$, Equation 1 yielded $p - .23$. Thus for the data on Table 5, the chi-square statistic is

$$\chi^2 = \frac{(35 - 50(.23) - .5)^2}{50(.23)} + \frac{(15 - 50(.77) - .5)^2}{50(.77)} = 63.83$$

(where 63.83 is significant beyond the .01 level). One would then be justified in assuming that the raters are exercising discriminatory power and the agreement is not due to chance.

TABLE 5

RATINGS OF 50 SEGMENTS OBTAINED FROM THREE RATERS, USING A WARMTH SCALE OF FIVE POINTS

| Segments | Rater | | | Segments | Rater | | |
|---|---|---|---|---|---|---|---|
| | A | B | C | | A | B | C |
| 1[a] | 5 | 4 | 5 | 26[a] | 1 | 1 | 2 |
| 2 | 5 | 2 | 3 | 27 | 5 | 4 | 2 |
| 3[a] | 1 | 1 | 2 | 28[a] | 1 | 1 | 1 |
| 4[a] | 4 | 4 | 4 | 29[a] | 2 | 2 | 1 |
| 5 | 4 | 3 | 2 | 30[a] | 3 | 2 | 2 |
| 6[a] | 4 | 4 | 3 | 31 | 2 | 4 | 5 |
| 7[a] | 5 | 5 | 5 | 32[a] | 5 | 4 | 5 |
| 8[a] | 4 | 3 | 4 | 33[a] | 2 | 2 | 2 |
| 9[a] | 5 | 5 | 4 | 34 | 3 | 2 | 1 |
| 10 | 2 | 3 | 2 | 35[a] | 1 | 2 | 3 |
| 11 | 4 | 3 | 2 | 36[a] | 3 | 3 | 2 |
| 12[a] | 5 | 5 | 4 | 37[a] | 3 | 3 | 2 |
| 13[a] | 3 | 3 | 4 | 38[a] | 4 | 3 | 4 |
| 14[a] | 5 | 5 | 4 | 39 | 5 | 4 | 2 |
| 15 | 3 | 2 | 1 | 40[a] | 1 | 2 | 1 |
| 16 | 1 | 2 | 3 | 41 | 2 | 4 | 5 |
| 17[a] | 5 | 5 | 4 | 42[a] | 2 | 2 | 2 |
| 18[a] | 4 | 4 | 3 | 43[a] | 1 | 1 | 2 |
| 19[a] | 4 | 4 | 4 | 44[a] | 2 | 2 | 2 |
| 20 | 3 | 2 | 1 | 45[a] | 3 | 3 | 3 |
| 21[a] | 2 | 2 | 1 | 46 | 2 | 4 | 5 |
| 22[a] | 2 | 2 | 2 | 47[a] | 1 | 1 | 1 |
| 23 | 4 | 3 | 2 | 48 | 1 | 2 | 3 |
| 24[a] | 3 | 2 | 3 | 49[a] | 3 | 3 | 3 |
| 25[a] | 3 | 3 | 3 | 50[a] | 2 | 3 | 3 |

[a] Agreement on a criterion of $r = 1$.

The rectangular model employed in computing $p$ is often not applicable since there is a general tendency for raters to avoid terminal categories in many cases, especially those involving deviations from a norm. This is called the error of central tendency, and, when it is present, the range of choice is narrowed and the true probability of agreeing by chance is higher than $p$; however, $p$ is still useful as a lower limit for that unknown probability of agreeing by chance so that if $Np$ is larger than

$N_1$, one would suspect lack of agreement among the raters.

The utilization of both coefficients of reliability and agreement yields a method of disclosing the variance accounted for by measurement via raters or judgments. The reliability techniques are statistical indexes of raters' consistency to judge phenomena toward the same end of a scale, although those judgments may or may not be commensurate. On the other hand, the agreement technique is an attempt to indicate raters making identical judgments of levels.

## REFERENCES

Ebel, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, 16, 407–424.

Ferguson, G. A. *Statistical analysis in psychology and education*. New York: McGraw-Hill, 1959.

(Received September 22, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 21–27

# SOME CONSEQUENCES OF THE REACTION TO AN AVERSIVE EVENT

RUSSELL M. CHURCH[1] AND DAVID J. GETTY

*Brown University*

In a free-responding situation an animal that merely reacts to an aversive event may appear to anticipate that event. To illustrate this problem, one of the typical reactions of a rat to a brief electric shock is considered. Although electric shock usually reduces the overall response rate, a given shock often leads to a burst of responses. The simulated behavior of such purely reactive subjects resembles the actual behavior of rats under conditions of nondiscriminative avoidance and punishment. Similar problems arise in other learning situations, and the yoked design is shown to be an inappropriate control, even without the assumption of individual differences. Several tests are described that distinguish the anticipation of an event from the reaction to an event.

In a free-responding situation an animal that merely reacts to an aversive event may appear to anticipate that event. No learning is necessary for an animal to react to an aversive event. Nonetheless, a purely reactive animal may appear to have learned the relationship between external stimuli or responses and the occurrence of an aversive event. The purpose of this article is to demonstrate the existence of the problem and to offer some suggestions for avoiding it.

## *An Example: A Reaction to an Aversive Event*

A rat in a lever box often makes a burst of responses following a brief electric shock. For example, if a rat is responding on a Sidman avoidance schedule in which each response postpones the occurrence of a train of shock pulses, the subject may make a burst of responses following each shock (Boren, 1961; Sidman, 1958).

Although such postshock bursts of responses are most familiar when an animal is responding on a Sidman avoidance schedule, similar reactions to an aversive event can occur on an appetitive base line. For example, Skinner (1938, p. 151) reported that the initial effect of a punishing slap of a lever in extinction was to decrease the latency to the next response. More recently, Weiss and Strongman (1969) presented brief shocks independently of responses during a warning signal to rats that had a history of positively reinforced lever pressing, and they observed postshock bursts of responding as well as a general suppression of appetitive lever pressing. Church, Wooten, and Matthews (1970a) trained rats to press a lever for variable-interval food reinforcement and then presented shocks on a random schedule. Figure 1 shows the mean response rate as a function of time since last shock of one of the rats in that experiment. Electric shock reduced the overall response rate, but a given shock often led to a flurry of responses.

For illustrative purposes, we will assume that responses are generated by a Poisson process with a rate parameter that is a function of time (a nonhomogeneous Poisson process). The rate parameter will be assumed to be an exponential function, $\gamma\ (t) = ae^{-bt} + c$, where $t$ is the time since last shock (in seconds), and $a$, $b$, and $c$ are constants. The mean response rate at $t = 0$ is $a + c$, where $c$ is the mean asymptotic rate. The rate of change in response rate is determined by the parameter $b$. In the subsequent discussion of Sidman avoidance, the values $a = 1$, $b = .3$, and $c = 2$ are used. These values provide a reasonable description of the reaction to shock of the subject shown in Figure 1.

The interpretation of such postshock responding is still a matter of conjecture. The electric shock may elicit some stereotyped response (Ulrich & Azrin, 1962) or a species-specific defense reaction (Bolles, 1970). The postshock bursts of responses may involve

---

[1] Requests for reprints should be sent to Russell M. Church, Walter S. Hunter Laboratory of Psychology, Brown University, Providence, Rhode Island 02912.

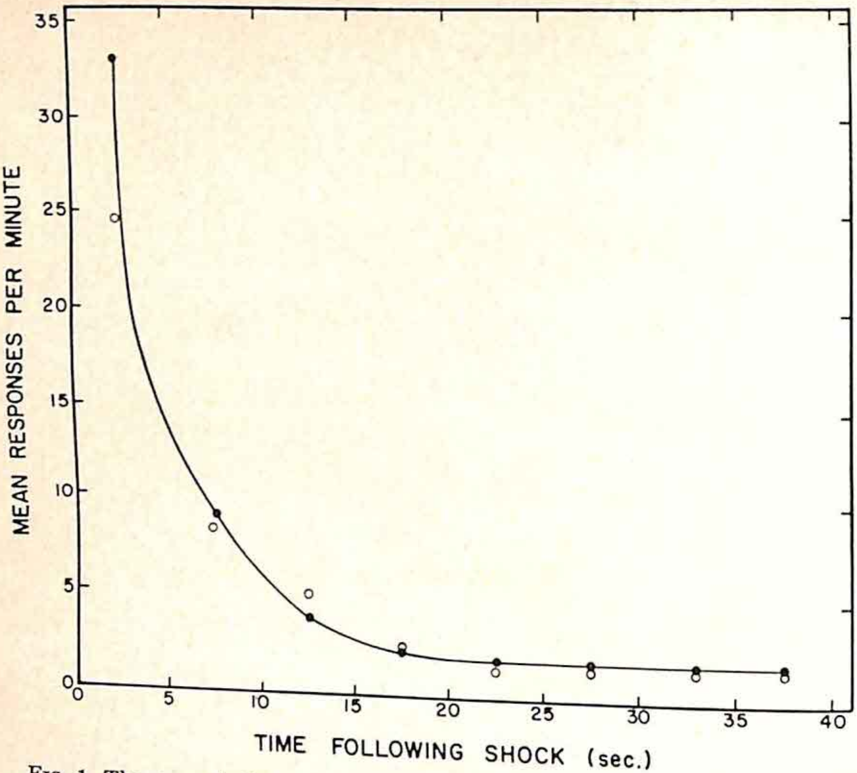


FIG. 1. The open circles represent response rate as a function of time following shock for one subject (Church, Wooten, & Matthews, 1970a). The solid circles represent the response rate during successive 5-second intervals if the instantaneous response rate were $e^{-.3t} + 2$, with $t$ = time following shock in seconds.

some form of reinforcement (e.g., delay the onset of the next shock in Sidman avoidance), or the shock may serve as a safety signal (Davis, 1970). Alternatively, if a response has been inhibited, the occurrence of a shock or any other intense stimulus may be sufficient to disinhibit the response. Some of these explanations can be excluded in specific instances, but the various alternative explanations of postshock responding have not been seriously tested. In any case, the present concern is with the consequences of the postshock bursts, not their cause.

*Avoidance*

Under conditions of nondiscriminative avoidance, each response of an animal postpones the occurrence of an aversive event. Normally electric shock is employed as the aversive event, and the number of seconds that a response postpones the shock is called the response–shock interval. (The number of seconds between shocks in the absence of a response is called the shock–shock interval.) The response–shock interval has a strong and reliable influence on the response rate (Sidman, 1953). With a shock–shock interval of 2.5 seconds, the response rate is a decreasing function of the response–shock interval between 2.5 and 150 seconds.

If the animal acquired a temporal discrimination it might respond shortly before a shock were due. Such an animal would respond more rapidly if a response postponed the next shock by only a few seconds than if it postponed the next shock for a longer period of time. A temporal discrimination, however, is not necessary for this result. An animal that merely reacted to an aversive event would also have a response rate that was inversely related to the response–shock interval.

Consider a subject in a Sidman avoidance experiment that emits a burst of responses whenever a shock occurs with a mean response rate that can be represented by the function shown in Figure 1 [$\gamma(t) = e^{-.3t} +$

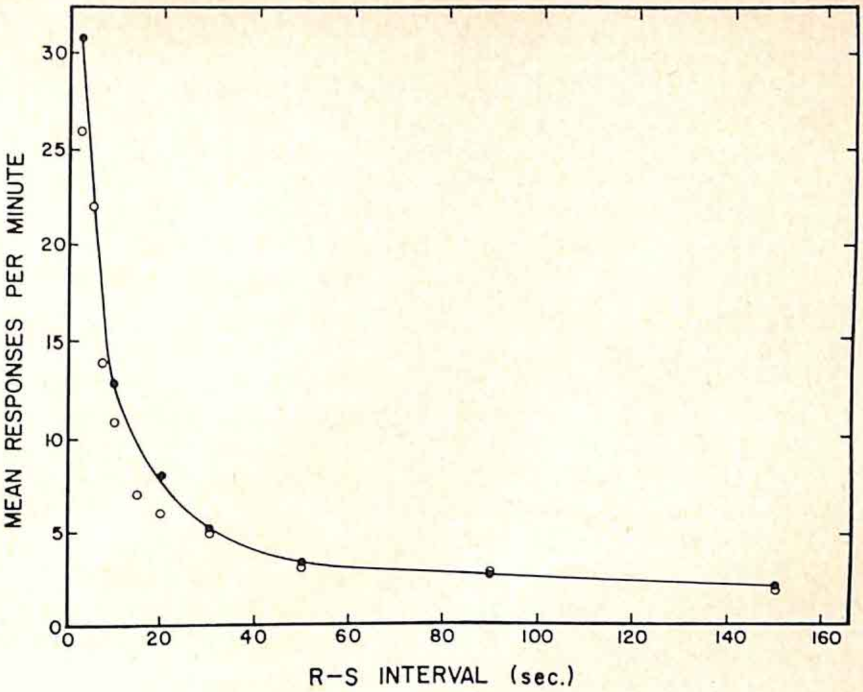


FIG. 2. The open circles represent response rate as a function of response-shock interval for the median of three subjects in a nondiscriminative avoidance procedure (Sidman, 1953). The solid circles represent the response rate as a function of the response–shock interval if the instantaneous response rate were $e^{-.3t} + 2$, with $t =$ time following shock in seconds.

2]. The responses of such a subject were simulated by computer according to the assumed Poisson process for 1-hour sessions under each of seven response–shock intervals. (The shock–shock interval was 2.5 seconds.) The probability of a response was calculated in each .2-second interval, and Figure 2 shows the mean response rate as a function of the response–shock interval. As the response–shock interval increases, the number of shocks received by a purely reactive subject decreases. The reduction in the number of shocks is reflected in the response rate (i.e., there is a reduction in the number of bursts of responses). The simulated results, based upon a subject that merely reacted to each shock, closely approximated the actual performance of the median of three subjects previously reported (Sidman, 1953).

The response rate of animals in a nondiscriminative avoidance situation is inversely related to the shock–shock interval, and additional noncontingent shocks lead to an increase in response rate. A burst of responses following shocks would produce both of these results. Of course, not all the behavior of subjects in a nondiscriminative avoidance situation can be explained on the basis of this model. For example, it does not account for the decrease in response rate when some of the responses are punished or the gradual course of extinction. Furthermore, there are many examples of Sidman avoidance in which the distribution of interresponse times gives evidence that the subject has acquired a temporal discrimination. Thus not all behavior in Sidman avoidance is simply a reaction to shock. The problem is to separate the behavior that is reactive from the behavior that is anticipatory.

### *Punishment*

The instrumental performance of an animal may be disrupted either by punishment or by noncontingent shock. The direct comparison of the response suppression produced by these two procedures has been assumed to provide a measure of the strength of association between the response and the aversive event. Under conditions of noncontingent shock the

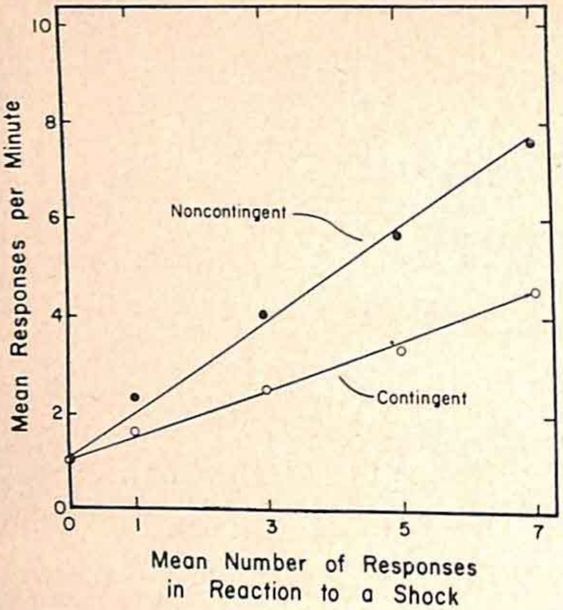


FIG. 3. Response rate of subjects with noncontingent shock and punishment (response-contingent shock) as a function of the mean number of responses in reaction to a shock. (The instantaneous response rate was $ae^{-.3t}+1$, with $t=$ time following shock in seconds.)

subject can only form an association between shock and situational cues; under conditions of punishment the subject can form an association between shock and response, as well as between shock and situational cues. Therefore, the magnitude of the difference between the response suppression produced by punishment and that produced by noncontingent shock has been considered to provide a measure of the strength of association between response and shock.

Under some conditions the degree of suppression produced by punishment and noncontingent aversive events is similar (Estes, 1944). Presumably, in these cases the suppression of the response is a consequence of association of shock with situational cues, not with the response. In most cases, however, punishment of a response has a far greater effect on that response than an aversive event that is independent of a response (e.g., Church, 1969; Rachlin & Herrnstein, 1969).

A typical procedure for the direct comparison of punishment with noncontingent shock is as follows: Two groups of rats are trained to press a lever for food reinforcement on a 1-minute variable-interval schedule of food reinforcement. After five 60-minute sessions of such training, the shock condition is introduced. During the next 10 sessions, when the 1-minute variable-interval schedule of food reward remains in effect, a shock is primed on a 1-minute random schedule. Under the noncontingent conditions the shock is delivered as soon as it is primed; under punishment conditions it is delivered immediately following the first response after it is primed. If it is assumed that a response-contingent and a noncontingent shock are equally effective, and each of them produces a response burst, it can be shown that the expected response rate of a subject would be lower in the contingent condition than in the noncontingent condition.

For a subject in the contingent condition there is a delay between the priming of a shock and its delivery following the next response. Any additional shocks primed during this delay period are not delivered. For a subject in the noncontingent condition, there is no delay between the priming and delivery of a shock, so the subject receives all of the shocks that are primed. This results in a lower expected shock rate during the response-contingent periods than the noncontingent periods. Since each shock is assumed to be followed by a response burst, the response rate of purely reactive subjects is expected to be lower during the contingent condition than during the noncontingent condition.

The responses of such subjects were simulated for a 2-hour session according to the assumed Poisson process, $\gamma(t) = ae^{-.3t} + 1$, under conditions of response-contingent and noncontingent shock. The probability of a response was calculated in each .2-second interval, and the results are shown in Figure 3. Despite the fact that the subjects never anticipated a shock but merely reacted to it, the response-contingent shock led to a slower response rate than the noncontingent shock. The values of the parameter $a$ were selected to provide the number of postshock responses indicated on the abscissa. The greater the reaction to shock, the greater the difference in response rate between punishment and noncontingent shock. (The fitted lines in Figure

3 are not meant to imply that the true function is linear.) The difference in response rate occurred even though the subjects did not learn an association between responses and shock.

*Failure of the Yoked Control Design*

The identification of a nonassociative effect of contingency has underscored the need to control shock frequency in the study of response-shock contingency. In an attempt to vary the contingency between a response and an aversive event, but hold constant the number and temporal distribution of shocks, some researchers have used the yoked control design. In this design an experimental subject receives a shock immediately following some of its responses, and a yoked control subject receives a shock at exactly the same time. Thus, the control subject receives the same number of shocks as the experimental subject with which it was yoked, but these shocks are uncorrelated with its responses.

Unfortunately, various sources of random error, such as individual differences in the effectiveness of the shock, lead to systematic differences between an experimental group and its yoked control group (Church, 1964). Some experimenters have sought to minimize the systematic effects of individual differences in the yoked control design by matching subjects prior to the experimental treatment. This is, at best, a partial solution, since the matching can never be perfect. In the next paragraph it is shown that even if the subjects are perfectly matched and have a constant value of all parameters throughout the experiment the yoked control design can produce spurious results.

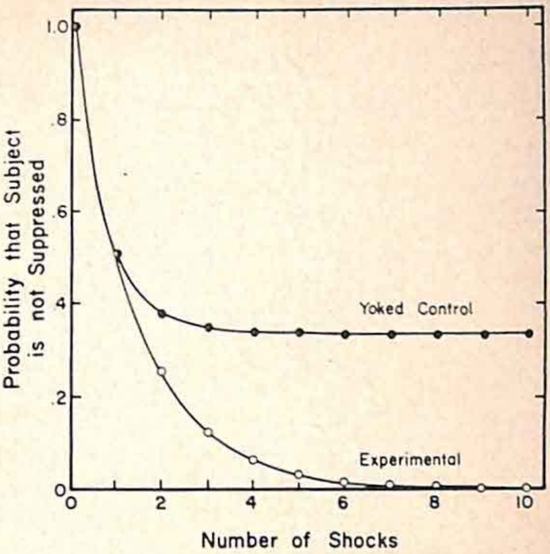


FIG. 4. Probability that a subject in the experimental group and its yoked control subject are not suppressed as a function of successive shocks. (The probability that any given shock will produce permanent suppression was set at .5 for both groups.)

Let us assume that a response-contingent shock and a noncontingent shock are equally effective in producing response suppression, that suppression is an all-or-none affair, and that there is no recovery from suppression. The probability that a given shock will lead to suppression is $p$, and the probability that it will not lead to suppression is $q = 1 - p$. After any given aversive event, an experimental subject with response-contingent shock will either be in a suppressed state (S) or a nonsuppressed state (N). Similarly, the yoked control subject with noncontingent shock will either be in state S or N. Thus the pair of subjects will be in one of four states after the $k$th shock (SS, SN, NS, or NN), with the first letter referring to the state of the experimental subject and the second letter referring to the state of the control subject. This defines a Markov chain with two transient states (NN and NS) and two absorbing states (SS and SN), for $q \neq 1$.

If both experimental and yoked control subjects begin in the nonsuppressed state, the probability that the response of an experimental subject is not suppressed following the $k$th shock is $q^k$, and the probability that the response of a control subject is not suppressed following the $k$th shock is $(q^{2k} + q)/(1 + q)$. Following the first shock, the probabilities that the experimental and yoked control subjects are not suppressed are both equal to $q$, but the two groups diverge with additional shocks. Figure 4 shows the probability that a subject is not suppressed as a function of trials with $p = .5$. The control subjects rapidly approach an asymptote of .33. When $k$ is sufficiently large, the probability that an experimental subject will not be suppressed approaches zero, while the probability that a yoked control subject will not be suppressed approaches $q/(1 + q)$. Therefore, the yoked control design cannot be used in the study of contingency between responses and aversive events.

### *Generality of the Problem*

The two major examples have been taken from instrumental training conditions, avoidance and punishment. The same problem arises in classical conditioning. For example, if an animal is responding for food on a variable-interval schedule of reinforcement and shocks are introduced at the end of each 1-minute interval, the animal might respond somewhat more slowly shortly before the shock was due than at other times. Although this could be evidence for temporal conditioning, it is also consistent with a burst of responses as a reaction to shock. Compensatory recovery in the conditioned emotional response (Estes, 1969) may also be a result of a reaction to shock. In other words, the response rate may increase following the shock at the end of a signal not to compensate for a low response rate during the signal but in reaction to the shock.

Although all of the examples have involved one particular reaction to electric shocks, other reactions to an aversive event would be equally difficult to separate from learned associations. Furthermore, there is no need to restrict attention to reactions to aversive events. Similar problems would arise with positive reinforcement if the response rate were directly influenced by the event. To conclude that an animal has learned a relationship between an external stimulus or a response and an aversive event, it is necessary to exclude the possibility that it is merely reacting to the aversive event. Various tests are available. For example, the reaction to shock cannot explain the results of a test during which no aversive events occur. Thus tests of the magnitude of the conditioned emotional response during extinction are appropriate (Rescorla, 1968). The concurrent chain procedure also separates the reaction to shock from a learned anticipation since the choice responses occur in the absence of the aversive event (Schuster & Rachlin, 1968). An alternative approach is to continue to use the aversive events during the test but to change either the external stimuli or the response manipulanda (Church, Wooten, & Matthews, 1970b). Such transfer tests separate the effects of the aversive events per se from the specific association between a stimulus or response and an aversive event. Finally, discrete-trial procedures may be used to introduce a long time interval between an aversive event and the next opportunity to respond.

Although there is sufficient evidence available to demonstrate most of the psychological phenomena described in this article, much of the supporting evidence has failed to distinguish between a learned anticipation of an event and a reaction to a previous event.

## REFERENCES

Bolles, R. C. Species-specific defense reactions and avoidance learning. *Psychological Review*, 1970, **77**, 32–48.

Boren, J. J. Isolation of post-shock responding in a free operant avoidance procedure. *Psychological Reports*, 1961, **9**, 265–266.

Church, R. M. Response suppression. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts, 1969.

Church, R. M. Systematic effect of random error in the yoked control design. *Psychological Bulletin*, 1964, **62**, 122–131.

Church, R. M., Wooten, C. L., & Matthews, T. J. Contingency between a response and an aversive event in the rat. *Journal of Comparative and Physiological Psychology*, 1970, **72**, 476–485. (a)

Church, R. M., Wooten, C. L., & Matthews, T. J. Discriminative punishment and the conditioned emotional response. *Learning and Motivation*, 1970, **1**, 1–17. (b)

Davis, H. Postshock responding on appetitive schedules: Aggression or discrimination? *Psychonomic Science*, 1970, **18**, 11–12.

Estes, W. K. An experimental study of punishment. *Psychological Monographs*, 1944, **57**(3, Whole No. 263).

Estes, W. K. Outline of a theory of punishment. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts, 1969.

Rachlin, H., & Herrnstein, R. J. Hedonism revisited: On the negative law of effect. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts, 1969.

Rescorla, R. A. Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 1–5.

Schuster, R., & Rachlin, H. Indifference between punishment and free shock: Evidence for the negative law of effect. *Journal of the Experimental Analysis of Behavior*, 1968, **11**, 777–786.

Skinner, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.

Sidman, M. Two temporal parameters of the maintenance of avoidance behavior by the white rat. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 253–261.

Sidman, M. Some notes on "bursts" in free-operant avoidance experiments. *Journal of the Experimental Analysis of Behavior*, 1958, **1**, 167–172.

Ulrich, R. E., & Azrin, N. H. Reflexive fighting in response to aversive stimulation. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 511–520.

Weiss, K. M., & Strongman, K. T. Shock-induced response bursts and suppression. *Psychonomic Science*, 1969, **15**, 238–240.

(Received October 7, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 28–31

# CRITIQUE OF "DESENSITIZATION: SOCIAL AND COGNITIVE FACTORS UNDERLYING THE EFFECTIVENESS OF WOLPE'S PROCEDURE"

GERALD C. DAVISON[1] AND G. TERENCE WILSON[2]

*State University of New York at Stony Brook*

An earlier study by Wilkins purported to demonstrate the overriding importance of various social and cognitive factors in systematic desensitization. The supporting evidence he cited is deficient on several counts, both logical and empirical.

Wilkins (1971) has concluded that research evidence demonstrates that the effectiveness of systematic desensitization is due to

> social variables involved in the patient-therapist relationship and cognitive variables involving expectancy of therapeutic gain, information feedback of success, training in the control of attention, and vicarious learning of the contingencies of behavior through instructed imagination [p. 311].

However, serious problems of both logic and data invalidate these several lines of argument and call this statement into serious question.

Wilkins began by arguing that neither hierarchy construction nor the pairing of graded aversives scenes with relaxation is a necessary element of successful systematic desensitization. The cognitive element of "instructed visualization" of fearful scenes is held to be sufficient. Unfortunately, the evidence Wilkins cited in support of this declaration consists of uncontrolled case studies (e.g., Wilson & Smith, 1968; Wolpin[3]), methodologically inadequate experiments (e.g., Cooke, 1966; Emery & Krumboltz, 1967; Wolpin & Raines, 1966; see also critiques by Bandura [1969] and Paul [1969]), or theoretical papers irrelevant to his argument (e.g., Davison, 1966; Rachman, 1968).

Even assuming that relaxation training and hierarchy construction are unnecessary procedural steps (a position suggested by Bandura's [1969] thorough review of the experimental literature), all that would be demonstrated is that Wolpe's (1958) theory of reciprocal inhibition is untenable. Wilkins (1971) implied that Rachman (1967) presented a generally agreed upon "learning theory" explanation in the form of the reciprocal inhibition hypothesis, "the appealing theoretical rationale" which is to be contrasted to social-cognitive interpretations. Yet Wilson and Davison (1971) have exposed the empirical and logical shortcomings of Wolpe's theory and, with Nelson (1966), have proposed a conditioning formulation more consistent with available experimental data. If systematic desensitization is construed as maximally facilitating nonreinforced exposure to the aversive stimulus (conditioned stimulus), then relaxation and hierarchy construction may be usefully seen as tactical maneuvers toward achievement of this goal rather than as indispensable theoretical ingredients (cf. Lang, 1969). At any rate, Wilkins cannot forcefully argue for any specific position merely by exclusion of reciprocal inhibition theory.

What evidence does Wilkins marshall in support of those social and cognitive factors that he views as underlying the effectiveness of systematic desensitization? That "expectancy of gain" plays a role in the efficacy of systematic desensitization would not be disputed by anyone familiar with the placebo literature. However, Wilkins ignored evidence clearly showing that by no means can all of the effect of systematic desensitization be attributed to this variable (cf. Davison, 1968; Lang, Lazovik, & Reynolds, 1965; Paul, 1966). Turning to the role of the therapist as a "social reinforcing agent" in systematic desensitization, we must point out that the Wagner and Cauthen (1968) study cited by Wilkins failed to show the relevance of *contingencies* of the alleged social reinforcement because of the absence of appropriate controls. Nor is statistical treatment of the sparse data provided. The study by Leitenberg, Agras, Barlow, and Oliveau (1969) showed only that strong placebo suggestions and praise for progress improve materially the efficacy of systematic desensitization, but, as the authors themselves advised, no conclusion can be drawn with respect only to social reinforcement. A second study in their series (Oliveau, Agras, Leitenberg, Moore, & Wright, 1969) was designed to partial out the placebo and social reinforcement factors; the findings revealed an impact only of the "therapeutic instructions" (placebo suggestions), not social reinforcement for progress. We do not find it possible, therefore, to agree with Wilkins' (1971) conclusion that these studies lend support to the importance of the therapist as a social reinforcer. Indeed, Wilkins cited several studies in which contact with the therapist was intentionally minimized (e.g., Kahn & Baker, 1968; Lang, 1969) without nullifying the efficacy of systematic desensitization, concluding:

---

[1] Requests for reprints should be sent to Gerald C. Davison, Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11790.

[2] Now at the Psychological Clinic, Rutgers University.

[3] M. Wolpin. The application of modified conditioning procedures to clinical problems: Three case histories. Paper presented at Camarillo State Hospital, Camarillo, California, 1966.

> These variables [expectancy of gain and social reinforcement from the therapist] in the therapeutic relationship appear to be quite powerful in that they can have an immediate and long-term effect following very minimal contact with the therapist [p. 313].

We fail to see the cogency of this conclusion, based as it is on data that show the effectiveness of systematic desensitization under conditions of little if any direct social contact with another human being, at least during the systematic desensitization procedure itself.

Wilkins (1971) confused several other issues. The information feedback used by Leitenberg, Agras, Thompson, and Wright (1968) was veridical, while in the study by Valins and Ray (1967) it was intentionally bogus. Wilkins failed to appreciate that the conceptual schemes of these two studies were fundamentally different. His consideration of the subject's perception of his own behavior as information feedback (cf. Bem, 1967) is interesting, yet it fails to come to grips with all the conditions that would seem to contribute to the subject's signaling anxiety. Bem's (1967) argument is that the subject makes an inference about his own attitude on the basis of his overt behavior, which another person could readily observe; thus, if one is not escaping from a shock, then that shock is not very painful. But in extrapolating from the Bandler, Madaras, and Bem (1968) study to systematic desensitization, Wilkins did not consider the instructions under which a systematic desensitization subject is operating, that is, to signal when feeling anxious. For the anxiety signal to be used as "information which can influence a subject's self-report of fear," it would seem necessary, as in the Bandler et al. (1968) study, to instruct one group of subjects to signal (i.e., escape from aversive scenes) and another not to signal *even when anxiety is felt.*

Perceived progress up the imaginal hierarchy was proposed by Wilkins (1971) as another information feedback parameter, but all the studies he cited showing a correlation of number of items passed with posttreatment improvement are readily interpreted in straight conditioning terms, as was done by the original investigators. A better test of this notion would entail bogus feedback on progress up the hierarchy, analogous to the manner in which Valins and Ray (1967) attempted to pit their cognitive interpretation of systematic desensitization against a conditioning interpretation.[4]

The third source of information feedback postulated by Wilkins (1971) centers on the social psychological work arising from Schachter's (1964) theory of emotion. Thus Wilkins endorsed the notion that phobics can be deceived into mislabeling physiological arousal as nonfearful (Valins & Ray, 1967). However, it must be remembered that unlike mildly fearful college students, phobic patients come for treatment with a background of extensive experience with, and discrimination learning in, their particular extreme fears that would probably lead to their rejecting the feedback as false. One of the studies Wilkins cited makes this very point:

---

[4] Brown [H. A. Brown. Systematic desensitization: Counterconditioning or expectancy manipulation? Unpublished doctoral dissertation, State University of New York at Stony Brook, 1970] has since attempted just such a strategy.

> Since we would not expect that an individual undergoing *extreme* pain, fear, or rage could easily be persuaded to attribute the accompanying physiological arousal to an artificial source, there should be limits placed on the generality of these notions. . . . Common sense would indicate that no amount of argument would persuade a man dodging machine-gun bullets that his physiological arousal was due to anything but the exigencies of the situation [Nisbett & Schachter, 1966, p. 228].

A corollary of Wilkins' emphasis on the importance of cognitive factors in Schachter's theory is his lack of due consideration to the physiological substrate of emotion (cf. Schachter & Wheeler, 1962). This error is further reflected in the omission of recent psychophysiological process research that clearly demonstrates that autonomic change during systematic desensitization is intimately tied to therapeutic success (cf. Lang, Melamed, & Hart, 1970).

Wilkins (1971) then asserted that "controlled attention shifts" are important in systematic desensitization: the subject being desensitized is said to become skilled at terminating his attention to the threatening situations, and this control over exposure, in itself, allegedly reduces fear. While Wilkins may be correct in affirming the importance of the patient's control over exposure to aversive scenes, it is unclear what explanatory value the concept of controlled attention shifts has per se. Phobics are certainly skilled at avoiding things they fear, and anyone who has tried to desensitize a severely phobic individual is familiar with the difficulties in inducing the person to *keep* his attention on a hierarchy item, whether imaginal or real. Furthermore, shifting one's attention from an aversive scene to the task of relaxing might be seen as an avoidance response that *prevents* the extinction of fear.

"Cognitive" conceptualizations are often put forward as alternatives to traditional conditioning formulations where in fact the postulated difference is only semantic. Thus, as a useful cognitive interpretation of systematic desensitization, Wilkins advanced vicarious learning by the client—by not avoiding the aversive scene, aversive consequences do not materialize. It is not clear what advantages accrue by restating in these Tolmanian terms the standard extinction formulation of nonreinforced exposure to the conditioned stimulus. Further, this is quite inconsistent with Wilkins' aforementioned hypothesis regarding "controlled attention shifts."

In summary, we feel that contrary to Wilkins' (1971) declarations, there is little if any research evidence for the notions he advances. Despite our negative comments on his article, however, we would hardly exclude the possibility that future research might demonstrate the importance of certain social/cognitive factors in systematic desensitization.

## REFERENCES

Bandler, R. J., Madaras, G. R., & Bem, D. J. Self-observation as a source of pain. *Journal of Personality and Social Psychology*, 1968, **9**, 205–209.

Bandura, A. *Principles of behavior modification*. New York: Holt, 1969.

Bem, D. Self perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 1967, **74**, 183–200.

Cooke, G. The efficacy of two desensitization procedures: An analogue study. *Behaviour Research and Therapy*, 1966, **4**, 17–24.

Davison, G. C. Anxiety under total curarization: Implications for the role of muscular relaxation in the desensitization of neurotic fears. *Journal of Nervous and Mental Disease*, 1966, **143**, 443–448.

Davison, G. C. Systematic desensitization as a counterconditioning process. *Journal of Abnormal Psychology*, 1968, **73**, 91–99.

Emery, J. R., & Krumboltz, J. O. Standard versus individualized hierarchies in desensitization to reduce test anxiety. *Journal of Counseling Psychology*, 1967, **14**, 204–209.

Kahn, M., & Baker, B. Desensitization with minimal therapist contact. *Journal of Abnormal Psychology*, 1968, **73**, 198–200.

Lang, P. J., Lazovik, A. D., & Reynolds, D. J. Desensitization, suggestibility, and pseudotherapy. *Journal of Abnormal Psychology*, 1965, **70**, 395–402.

Lang, P. J. The mechanics of desensitization and the laboratory study of fear. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969.

Lang, P. J., Melamed, B. G., & Hart, J. A psychophysiological analysis of fear modification using an automated desensitization procedure. *Journal of Abnormal Psychology*, 1970, **76**, 220–234.

LEITENBERG, H., AGRAS, W. S., BARLOW, D. H., & OLIVEAU, D. C. Contribution of selective positive reinforcement and therapeutic instructions to systematic desensitization therapy. *Journal of Abnormal Psychology,* 1969, **74**, 113–118.

LEITENBERG, H., AGRAS, W. S., THOMPSON, L. E., & WRIGHT, D. E. Feedback in behavior modification: An experimental analysis in two phobic cases. *Journal of Applied Behavior Analysis,* 1968, **1**, 131–137.

NELSON, F. Effects of two counterconditioning procedures on the extinction of fear. *Journal of Comparative and Physiological Psychology,* 1966, **62**, 208–213.

NISBETT, R. E., & SCHACHTER, S. The cognitive manipulation of pain. *Journal of Experimental Social Psychology,* 1966, **2**, 227–236.

OLIVEAU, D. C., AGRAS, W. S., LEITENBERG, H., MOORE, R. C., & WRIGHT, D. E. Systematic desensitization, therapeutically oriented instructions and selective positive reinforcement. *Behaviour Research and Therapy,* 1969, **7**, 27–33.

PAUL, G. L. *Insight vs. desensitization in psychotherapy: An experiment in anxiety reduction.* Stanford, Calif.: Stanford University Press, 1966.

PAUL, G. L. Outcome of systematic desensitization. II. Controlled investigations of individual treatment, technique variations, and current status. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status.* New York: McGraw-Hill, 1969.

RACHMAN, S. Systematic desensitization. *Psychological Bulletin,* 1967, **67**, 93–103.

RACHMAN, S. The role of muscular relaxation in desensitization therapy. *Behaviour Research and Therapy,* 1968, **6**, 159–166.

SCHACHTER, S. The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. I. New York: Academic Press, 1964.

SCHACHTER, S., & WHEELER, L. Epinephrine, chlorpromazine, and amusement. *Journal of Abnormal and Social Psychology,* 1962, **65**, 121–128.

VALINS, S., & RAY, A. A. Effects of cognitive desensitization on avoidance behavior. *Journal of Personality and Social Psychology,* 1967, **7**, 345–350.

WAGNER, M. K., & CAUTHEN, N. R. A comparison of reciprocal inhibition and operant conditioning in the systematic desensitization of a fear of snakes. *Behaviour Research and Therapy,* 1968, **6**, 225–227.

WILKINS, W. Desensitization: Social and cognitive factors underlying the effectiveness of Wolpe's procedure. *Psychological Bulletin,* 1971, **76**, 311–317.

WILSON, A., & SMITH, F. J. Counterconditioning therapy using free association: A pilot study. *Journal of Abnormal Psychology,* 1968, **73**, 474–478.

WILSON, G. T., & DAVISON, G. C. Processes of fear reduction in systematic desensitization: Animal studies. *Psychological Bulletin,* 1971, **76**, 1–14.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford, Calif.: Stanford University Press, 1958.

WOLPIN, M., & RAINES, J. Visual imagery, expected roles and extinction as possible factors in reducing fear and avoidance behavior. *Behaviour Research and Therapy,* 1966, **4**, 25–37.

(Received October 8, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 32–36

# DESENSITIZATION:
## GETTING IT TOGETHER WITH DAVISON AND WILSON

WALLACE WILKINS [1]

*University of Miami*

In spite of Davison and Wilson's global criticisms, the conclusions originally presented by Wilkins in his 1971 review are supported by empirical evidence. Neither hierarchy construction nor training in muscle relaxation, but only the instructed imagination of fear-relevant scenes, is a necessary element of Wolpe's systematic desensitization procedure. Factors influential in successful systematic desensitization include expectancy of therapeutic gain, reinforcing qualities of the therapist, feedback of success, and training in the control of attention. Aspects of the procedure are quite similar to and consistent with modeling procedures of fear reduction. Davison and Wilson appear to be in considerable agreement with several of these conclusions.

Davison and Wilson (1972) criticized Wilkins' (1971) review of social and cognitive factors contributing to successful systematic desensitization. The conclusions they have drawn indicate that their zeal has, in some cases, created a straw man, in other cases caused them to overlook problems in their own logic, and in still other cases led them to criticize positions taken by Wilkins (1971) with which Davison and Wilson are in basic agreement. Other criticisms leveled by Davison and Wilson are sound and, in the interest of identifying and understanding the many variables involved in systematic desensitization, are well received.

### *On the Inadequacy of Wolpe's Theory*

Davison and Wilson claimed that a bevy of evidence Wilkins (1971) presented (Cooke, 1966; Davison, 1966; Emery & Krumboltz, 1967; Rachman, 1968; Wilson & Smith, 1968; Wolpin, personal communication, 1966; Wolpin & Raines, 1966) did not warrant Wilkins' conclusion that neither hierarchy construction, nor training in muscle relaxation, nor the pairing of graded aversive scenes with muscle relaxation is a necessary element of successful systematic desensitization. The studies by Wilson and Smith (1968) and Wolpin (personal communication, 1966), as uncontrolled as they may be, do offer clear demonstrations that desensitization can occur without including all the components of Wolpe's (1958) procedure. Logically, then, these components may not be necessary. Further demonstrations by Cooke (1966), Wolpin and Raines (1966), and Lazarus (1965) that fear reduction can occur in the presence of muscle tension offer additional suggestive evidence that muscle relaxation may not be necessary.

The suggestive evidence offered by these uncontrolled case studies is validated by experiments inappropriately criticized by Davison and Wilson (1972). Davison and Wilson criticized studies by Cooke (1966), Emery and Krumboltz (1967), and Wolpin and Raines (1966) as methodologically inadequate, and referred the reader to Paul (1969) and Bandura (1969) for critiques.

However, neither Bandura (1969) nor Paul (1969) offered the criticisms of these studies that Davison and Wilson lead the reader to believe. Bandura (1969) did not criticize two of the studies and briefly commented on the third (Wolpin & Raines, 1966) about a point totally irrelevant to the question of the necessity of the components of systematic desensitization. It is evident from Paul's (1969) evaluative analyses of the Cooke (1966) and Emery and Krumboltz (1967) studies that the soundness of the methodologies employed do, indeed, warrant Paul's own conclusion that cause–effect relationships between treatment and improvement were adequately demonstrated. Paul (1969) added that since data were not provided for differential therapist effectiveness, it was impossible to tell how much improvement was due to systematic desensitization per se and how much was due to therapist characteristics. However, this issue is independent of the issue concerning the necessity of various components of systematic desenstization: if the data, in fact, show therapist effects to be minimal, these studies still indicate that individualized hierarchies and muscle relaxation are not necessary to achieve significant fear reduction. On the other hand, data showing that the entire treatment effect was due to therapist variables would also be evidence that the elements of the systematic desensitization procedure are not necessary for fear reduction. Davison and Wilson's citation of Bandura (1969) and Paul (1969) as criticism of Cooke (1966), Emery and Krumboltz (1967), and Wolpin and Raines (1966) is clearly misleading.

---

[1] Requests for reprints should be sent to Wallace Wilkins, Department of Psychology, University of Miami, Coral Gables, Florida 33124.

Studies cited by Wilkins (1971) that Davison and Wilson (1972) have not labeled as "uncontrolled, methodologically inadequate, or irrelevant" (Garfield, Darwin, Singer, & McBrearty, 1967; Leitenberg, Agras, Barlow, & Oliveau, 1969; Ritter, 1968) are further support for the conclusion that hierarchy construction, muscle relaxation training, and their pairing are not necessary conditions.

In spite of Davison and Wilson's (1972) protestations of "irrelevance," the issues explored in Davison (1966) and Rachman (1968) are extremely cogent to a core issue in Wilkins' (1971) review—"that Wolpe's theoretical formulations are inadequate in identifying the critical variables involved in treatment outcome [p. 312]."

Ironically, Wilson and Davison (1972) agreed with Wilkins (1971) by stating "all that would be demonstrated is that Wolpe's (1958) theory of reciprocal inhibition is untenable [p. 28]." This was basically what Wilkins intended to demonstrate in the early portion of his review.

Wilson and Davison (1972) claimed that Wilkins (1971) presented the above studies as evidence for the declaration that "instructed visualization of fearful scenes is held to be sufficient [Wilson & Davison, 1972, p. 28]." In the first place, Wilkins did not declare that instructed visualization was a *sufficient* condition for desensitization, only that "of the elements included in Wolpe's procedure, instructed visualization of hierarchy scenes appears to be the only one *necessary* [italics added] for treatment success [Wilkins, 1971, p. 312]." Only two of the above studies were cited (Wolpin & Raines, 1966; Wolpin, personal communication, 1966) in the tentative statement that instructed visualization *may* be a sufficient condition.

In their next "argument" against Wilkins (1971), Davison and Wilson (1972) agreed with Wilkins. By stating that relaxation and hierarchy construction may be usefully seen as tactical maneuvers rather than as "indispensable theoretical ingredients," they are in agreement with Wilkins' conclusion that Wolpe's theoretical formulations of systematic desensitization are inadequate. In addition, their conceptualization of systematic desensitization as "maximally facilitating nonreinforced exposure to the aversive stimulus," while relaxation training and hierarchy construction are conceptualized as only tactical maneuvers, is consistent with Wilkins' (1971) conclusion that neither hierarchy construction nor training in muscle relaxation is a necessary component, but instructed visualization (exposure) is the only necessary element of the procedure introduced by Wolpe.

Indeed, if one reads behind the vendetta generated by Davison and Wilson (1972), it appears that Wilkins (1971) and Davison and Wilson (1972) are in substantial agreement over the first 96 words of Wilkins' (1971) 136-word abstract.

### *On Expectancy of Therapeutic Gain*

Turning to social and cognitive variables that Wilkins (1971) presented as influential in the outcome of systematic desensitization, Davison and Wilson (1972) appeared to agree that expectancy of therapeutic gain is a prominent variable and then, in a strawman gesture, criticized Wilkins' position by citing literature that shows "that by no means can *all* of the effects of systematic desensitization be attributed to this variable [pp. 28–29]." Wilkins made no such claim. On the level of elementary logic, Wilkins could not have been misconstrued as endorsing such a position, for that would be equivalent to stating that "expectancy" is a sufficient condition and that no other condition is necessary; but Wilkins had stated previously that instructed imagination of scenes is a necessary condition.

### *On the Therapist as a Reinforcer of Nonfear Behavior*

Davison and Wilson's (1972) criticisms of Wilkins' (1971) conclusions regarding the effect of the therapist as a social reinforcing agent are quite appropriate. The conclusion Wilkins reached on this issue extended well beyond the data reviewed. Although the data included in Wilkins' review are not adequate support, the conclusions do appear valid in light of more recent studies. Agras, Leitenberg, and Barlow (1968) and Agras, Leitenberg, Barlow, and Thompson (1969), by reversing contingencies of verbal reinforcement delivered by the therapist, demonstrated that this variable has a potent effect on the reduction of phobic behavior.

### *On the Role of Information Feedback*

Wilkins presented studies that demonstrated that both veridical (Leitenberg et al., 1968) and bogus but credible (Valins & Ray, 1967) feedback of progress facilitated fear reduction. Wilkins did not "confuse the issues" (as Davison and Wilson claimed) by presenting these two empirical studies. Wilkins does "appreciate that the conceptual schemes of these two studies were fundamentally different [Davison & Wilson, 1972, p. 29];" however, it is not confusing that both procedures did, indeed, facilitate fear reduction.

Given that information feedback does influence fear reduction, Wilkins (1971) was then criticized for exploring, in public, "several *possible* sources of information feedback that *may* contribute [italics added; Wilkins, 1971, p. 313]" to fear reduction. Certainly, it is an empirical question as to whether the three possible sources of information explored by Wilkins actually do serve this facilitating function. Davison and Wilson constructively offered strategies for answering these empirical questions.

In another straw-man gesture, Davison and Wilson concluded that Wilkins "endorsed the notion that phobics can be deceived into mislabeling physiological arousal as nonfearful [Davison & Wilson, 1972, p. 29]." Wilkins (1971) did report a demonstration (Valins & Ray, 1967) whereby fear (not phobia) was reduced via bogus feedback, but Wilkins never made the endorsement that Davison and Wilson claimed. Where Davison and Wilson presented a common-sense notion as evidence against such an endorsement, Wilkins preferred to view it as an empirical question as to whether (and under what conditions) a procedure like Valins and Ray's (1967) may be effective in treating clinical phobias.

Davison and Wilson pointed out that Wilkins omitted recent research that indicates that physiological changes do occur through systematic desensitization. Davison and Wilson are correct. This evidence, almost all of which has been published since Wilkins' review was completed, has been reviewed by Mathews (1971). That physiological changes do occur does not preclude the possibility that therapeutic gain can occur through a relabeling of physiological arousal. Further research is needed to clarify (*a*) whether the physiological changes precede or follow attitudinal, emotional, or behavioral changes; (*b*) under what condition changes can occur in some of these fear modalities without accompanying changes in others.

### *On the Role of Instructed Imagination*

1. *Controlled attention shifts.* Wilkins (1971) asserted that imagination of aversive scenes is the only necessary element of Wolpe's procedure and that two aspects of this element appear to be important in fear reduction: controlled attention shifts and exposure to contingencies of nonavoidant behavior in the fear-producing situation. Contrary to the claims of Davison and Wilson (1972), Wilkins did not also assert that control over exposure, in itself, reduces fear. It is unlikely that shifting attention away from aversive stimuli is, in itself, therapeutic. Davison and Wilson correctly pointed out that "shifting one's attention from an aversive scene to the task of relaxing might be seen as an avoidance response that prevents the extinction of fear [Davison & Wilson, 1972, p. 30]." However, if such avoidance responses are reinforced by fear reduction, what main-

tains the client's behavior of returning for subsequent desensitization session, reexposing himself to the scenes, and progressing up the hierarchy? Could some of the factors maintaining this behavior possibly include expectancy of gain, reinforcing properties of the therapist, and feedback of success?

In addition to shifting attention from the aversive stimuli, the systematic desensitization procedure also includes controlled attention shifts toward the aversive stimuli. The imagination of fear-relevant scenes appears to be the only necessary component of Wolpe's procedure, and the shifting of attention toward these scenes appears to be a critical variable offering a great potential in integrating the social, cognitive, physiological, and conditioning aspects of systematic desensitization. Experimentation around this focal variable into (*a*) the factors that precede and maintain the activity of shifting attention toward aversive scenes and (*b*) the changes produced as a consequence of exposure to these stimuli may eventually produce a parsimonious social-cognitive-physiological-learning model of psychotherapy.

2. *Exposure to contingencies of nonavoidant behavior.* The functional importance of controlled attention shifts toward aversive stimuli can be accounted for in terms of theoretical explanations derived from both operant and respondent models. The respondent interpretation is that fear is extinguished because the imaginal conditioned stimuli are not followed by the unconditioned stimuli. An interpretation derived from the operant model is that contingencies of nonavoidant behavior are learned through the imaginal observation of the absence of negative reinforcers and/or the presence of positive reinforcers of nonavoidant behavior in the imagined situation. This imaginal contingency learning is hypothesized to be functionally similar to the observational learning common to the procedures of modeling (Bandura, Grusec, & Menlove, 1967; Geer & Turteltaub, 1967), contact desensitization (Ritter, 1968, 1969), and participant modeling (Rimm & Mahoney, 1969; Rimm & Medeiros, 1970).

The differences between the explanations derived from respondent and operant paradigms are, contrary to claims by Davison and Wilson (1972), more than semantic, and neither explanation is necessarily Tolmanian. That the scenes typically employed in systematic desensitization are scenes of the individual *behaving* in relation to the feared stimulus (e.g., picking up a spider, stepping on a snake) is consistent with a hypothesis involving fear reduction through modeling of imagined nonavoidant behavior; however, as Wilkins (1971) originally stated, no empirical evidence bearing directly on this speculation has yet been generated.

In spite of the critical evaluation by Davison and Wilson (1972), conclusions originally presented by Wilkins (1971) still appear to be substantiated. Neither hierarchy construction nor training in muscle relaxation is necessary; the instructed imagination of fear-relevant scenes is the only necessary element of Wolpe's (1958) procedure. Variables that research evidence indicates are influential in successful therapeutic outcome include expectancy of therapeutic gain, the social reinforcing qualities of the therapist, information feedback of success, and training in the control of attention. The modelinglike learning of contingencies of nonavoidant behavior in the fearful situation remains a speculation.

## REFERENCES

AGRAS, W. S., LEITENBERG, H., & BARLOW, D. H. Social reinforcement and the modification of agoraphobia. *Archives of General Psychiatry,* 1968, **19,** 423–427.

AGRAS, S., LEITENBERG, H., BARLOW, D. H., & THOMSON, L. E. Instructions and reinforcement in the modification of neurotic behavior. *American Journal of Psychiatry,* 1969, **125,** 1435–1439.

BANDURA, A. *Principles of behavior modification.* New York: Holt, 1969.

BANDURA, A., GRUSEC, J. E., & MENLOVE, F. L. Vicarious extinction of avoidance behavior. *Journal of Personality and Social Psychology,* 1967, **5,** 16–23.

COOKE, G. The efficacy of two desensitization procedures: An analogue study. *Behaviour Research and Therapy,* 1966, **4,** 17–24.

DAVISON, G. C. Anxiety under total curarization: Implications for the role of muscular relaxation in the desensitization of neurotic fears. *Journal of Nervous and Mental Disease,* 1966, **143,** 443–448.

DAVISON, G. C., & WILSON, G. T. Critique of "Desensitization: Social and cognitive factors underlying the effectiveness of Wolpe's procedure." *Psychological Bulletin,* 1972, **78,** 28–31.

Emery, J. R., & Krumboltz, J. D. Standard versus individualized hierarchies in desensitization to reduce test anxiety. *Journal of Counseling Psychology*, 1967, **14**, 204–209.

Garfield, Z. H., Darwin, P. L., Singer, B. A., & McBrearty, J. F. Effect of "in vivo" training of experimental desensitization of a phobia. *Psychological Reports*, 1967, **20**, 515–519.

Geer, J. H., & Turteltaub, A. Fear reduction following observation of a model. *Journal of Personality and Social Psychology*, 1967, **6**, 327–331.

Lazarus, A. A. A preliminary report on the use of directed muscular activity in counter-conditioning. *Behaviour Research and Therapy*, 1965, **2**, 301–303.

Leitenberg, H., Agras, W. S., Barlow, D. H., & Oliveau, D. C. Contribution of selective positive reinforcement and therapeutic instructions to systematic desensitization therapy. *Journal of Abnormal Psychology*, 1969, **74**, 113–118.

Leitenberg, H., Agras, W. S., Thompson, L. E., & Wright, D. E. Feedback in behavior modification: An experimental analysis in two phobic cases. *Journal of Applied Behavior Analysis*, 1968, **1**, 131–137.

Mathews, A. M. Psychophysiological approaches to the investigation of desensitization and related procedures. *Psychological Bulletin*, 1971, **76**, 73–91.

Paul, G. L. Outcome of systematic desensitization: II. Controlled investigations of individual treatment, technique variations, and current status. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969.

Rachman, S. The role of muscular relaxation in desensitization therapy. *Behaviour Research and Therapy*, 1968, **6**, 159–166.

Rimm, D. C., & Mahoney, M. J. The application of reinforcement and participant modeling procedures in the treatment of snake-phobic behavior. *Behaviour Research and Therapy*, 1969, **7**, 369–376.

Rimm, D. C., & Medeiros, D. C. The role of muscle relaxation in participant modeling. *Behaviour Research and Therapy*, 1970, **8**, 127–132.

Ritter, B. The group desensitization of children's snake phobias using vicarious and contact desensitization procedures. *Behaviour Research and Therapy*, 1968, **6**, 1–6.

Ritter, B. Treatment of acrophobia with contact desensitization. *Behaviour Research and Therapy*, 1969, **7**, 41–45.

Valins, S., & Ray, A. A. Effects of cognitive desensitization on avoidance behavior. *Journal of Personality and Social Psychology*, 1967, **7**, 345–350.

Wilkins, W. Desensitization: Social and cognitive factors underlying the effectiveness of Wolpe's procedure. *Psychological Bulletin*, 1971, **76**, 311–317.

Wilson, A., & Smith, F. J. Counterconditioning therapy using free association: A pilot study. *Journal of Abnormal Psychology*, 1968, **73**, 474–478.

Wolpe, J. *Psychotherapy by reciprocal inhibition*. Stanford, Calif.: Stanford University Press, 1958.

Wolpin, M., & Raines, J. Visual imagery, expected roles and extinction as possible factors in reducing fear and avoidance behavior. *Behaviour Research and Therapy*, 1966, **4**, 25–37.

(Received October 14, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 37–48

# NIGHTMARE BEHAVIOR:

## A REVIEW

MICHEL HERSEN[1]

*Veterans Administration Center, Jackson, Mississippi*

Objective reports, surveys, observations, and studies relating to nightmare behavior are reviewed. Physiological correlates of nightmares including the relationship of homeostatic imbalances and production of nightmares are examined. Investigations are criticized on the basis of inadequate numbers of subjects, absence of controls, excessive reliance on self-reports, poor reliability and validity, and lack of experimental control over independent variables under study. Possible improvements in existing methodology are considered, and suggestions for further research are offered. Psychogenic theories accounting for production of nightmare dreams are reviewed. In addition, recent applications of behavioral methodology to treatment of pathological nightmare behavior are surveyed.

The occurrence of nightmares appears to be a rather common phenomenon, both for normal and pathological populations, as documented by a survey conducted by Cason (1935) three and one-half decades ago. Similar findings have also been reported in more recent investigations (Feldman & Hersen, 1967; Hersen, 1971a). Feldman and Hersen (1967), examining a college student population, found that of 1,317 subjects filling out a brief questionnaire, 86% admitted to experiencing at least one nightmare dream in the year prior to being surveyed. In this study a nightmare is operationally defined as a frightening dream that awakens the dreamer.

Despite the apparent ubiquity of the nightmare experience, only recently have objective personality and physiological evaluations been made of the nightmare sufferer (Broughton, 1968; Feldman & Hersen, 1967; Fisher, Byrne, Edwards, & Kahn, 1969; Hersen, 1971a). By contrast, much theorizing as to etiology and course of the nightmare has appeared (Fenichel, 1945; Hadfield, 1954; Jones, 1949; de Verteuil, 1962). But most of the theoretical formulations proposed have not been empirically validated in a systematic fashion, especially as they generally are not amenable to experimental assessment. Only "clinical" proof in the form of case histories has been adduced in support of theoretical positions espoused (Bellamy, 1915; Jones, 1949; Pearson, 1949; Stern, 1951; de Verteuil, 1962.

In the present article the primary focus is on a critical review of the objective reports, surveys, observations, and studies relating to nightmare experiences. Physiological correlates of nightmares including the relationship of homeostatic imbalances and production of nightmares are also examined. A number of case histories are presented with particular emphasis on recent application of behavioral methodology to treatment of pathological nightmare behavior. In addition, some of the psychogenic theories that have been developed to account for manifestation of nightmare behavior are presented in summary form.

### SELF-REPORTED NIGHTMARE FREQUENCY

Cason (1935) conducted an extensive survey, published in monograph form, in which a variety of issues relating to incidence of nightmares was studied. Despite the fact that most subjects interviewed had experienced a nightmare in the past, the study was only concerned with subjects whose nightmares occurred in the month before being surveyed.

Cason (1935) reported the following percentage of subjects having nightmares on a one per month basis: normal children, 32%; summer session college subjects, 19%; blind students, 18%; retarded subjects, 13%; psychiatric subjects, 7%. Blanchard (1926) found that of 300 children under 18 years of age treated in a child guidance clinic, 13% had dreams that were classified under the heading of fear dreams; some of these were categorized as nightmares. Feldman and Hersen (1967) indicated that 29% of their college subjects reported experiencing nightmares once monthly, whereas Hersen (1971a) reported that 32% of his psychiatric sample had nightmares on a one per month basis. A possible explanation for the large discrepancies among the Cason (1935), Feldman and Hersen (1967), and Hersen (1971a) studies, other than mere sampling differences, is that Cason (1935) actually interviewed his subjects, whereas Feldman and Hersen (1967) and Hersen (1971a) simply required their subjects to fill out a brief questionnaire. It is therefore conceivable that the subjects in the face-to-face interview situation are more easily threatened and are less likely to admit to what they construe as representing weaknesses on their part.

---

[1] Also at the University of Mississippi Medical Center.

Requests for reprints should be sent to Michel Hersen, Psychology Service, Veterans Administration Center, Jackson, Mississippi 39216.

Cason (1935) found that blind and normal adults with nightmares required longer periods of time to fall asleep than their non-nightmare counterparts. Contrary to Jones' (1949) contention that the supine sleeping position was associated with occurrence of nightmares, Cason (1935) noted that his subjects reported the side as being the most frequent sleeping position at the inception of the nightmare. As might be expected, psychiatric inpatients listed the greatest number of peculiar sleeping positions when their nightmares began. However, these reports are obviously subject to retrospective distortion and are of little empirical value.

In all nightmare groups interviewed, 60%–90% of the subjects noted that they were awakened by their nightmares (Cason, 1935). When asked to describe the effects of nightmares on subsequent awakening, subjects' reports ranged from "no later influence" to "feel afraid," "remain awake," "feel nervous and excited," "feel weak, exhausted, or fatigued," and "be uncertain about the reality of the nightmare."

With regard to recurrence of nightmares, 31%–55% of Cason's (1935) subjects stated that they previously had experienced the same nightmare dream on at least one occasion. Feldman and Hersen (1967) found that recurrence of nightmare rates was linearly related to admitted nightmare frequency. The seldom group (once a year) had a rate of 3%, the sometimes group (once a month) a rate of 37%, and the frequent group (once a week) a rate of 84%.

In comparing rates of nightmares for different populations, it will be recalled that Feldman and Hersen (1967) found that 86% of their college subjects had experienced at least one nightmare dream in the past year. In a modified replication, Hersen (1971a) found that 57% of a hospitalized psychiatric population admitted having one nightmare in the previous year. Discrepancies in the unexpected direction between normal and pathological populations have also been noted by Cason (1935), but have been attributed to inadequate cooperation on the part of the mental patient. However, in Hersen's (1971a) investigation, the comparatively reduced proportion of psychiatric inpatient nightmare sufferers might be related to suppressive influences of psychotropic medication being administered. Zung (1970) pointed out that percentage of sleeping time in the rapid eye movement (REM) stage for psychiatric patients that are administered phenothiazines is reduced. In this connection it is also noted that Broughton (1968) linked nightmares, somnambulism, enuresis, and night terrors as sleep disturbances all preferentially occurring during *arousal* from slow-wave sleep as opposed to REM sleep.

In spite of the general ubiquity of the nightmare experience, the incidence of frequent nightmares in normal and pathological populations is small. Feldman and Hersen (1967), using undergraduates as subjects, found that only 5% admit to experiencing nightmares on at least a once a week basis. Hersen (1971a), assessing a psychiatric population, reported an incidence of 7% frequent nightmare sufferers. In both the Feldman and Hersen (1967) and Hersen (1971a) reports, female subjects admitted experiencing more frequent nightmares than their male counterparts. This finding, though not at all surprising, may reflect a greater fearfulness on the part of female subjects (Geer, 1965; Hersen, 1971b). In addition, it appears that a social desirability factor might be operating in that male subjects, in order to preserve their masculine image, are less likely to admit to their nightmare experiences. The aforementioned could be studied indirectly through the use of social desirability scales (Ford, 1964), or could be evaluated directly by comparing admitted frequency rates with observed frequency rates in an experimental sleep laboratory. However, in this kind of evaluation the decreased likelihood of nightmare dreams occurring in the experimental sleep laboratory would have to be considered (Hartmann, 1970).

### NIGHTMARE THEMES

It is of historical interest that thematic material elicited from reports of nightmare sufferers has altered markedly over the centuries. Jones (1949) and Gastaut and Broughton (1965), commenting on descriptions of nightmares from writings of the Middle Ages, noted the frequency with which witches, demons, devils, and other assorted unpleasant characters made their appearances in nightmare dreams. For example, Fuseli, an eighteenth century Anglo-Swiss painter, is renowned for his portrayal "The Nightmare" in which a short, squat, sinister-looking figure reposes on the sleeper's chest (Schneck, 1969). However, themes permeating twentieth century nightmares appear to represent more concern with attempts at resolving conflicts (Cason, 1935; Hadfield, 1954), enclosure in unpleasant surroundings (Gastaut & Broughton, 1965), death, injury, and social embarrassment (Feldman & Hyman, 1968).

According to Jones (1949) the nightmare dream leaves the hapless dreamer with intense feelings of agonizing dread, a conviction of helpless paralysis, and a sense of weight and oppression on the chest. Undoubtedly contributing to these subjective feelings are the generally terrifying themes permeating the majority of nightmare dreams. Cason (1935) categorized the most prominent subject matter presented by his nightmare subjects in order of their reported frequency: animals, death and murder, being chased, home and family, miscellaneous people, falling, and accidents.

Feldman and Hyman (1968) performed a content analysis of retrospective nightmare reports of 49 women and 21 men. The subjects' frequent nightmares were compared with a control series of nonnightmare dreams. By contrast to nightmare dreams, in "ordinary" dreams only less than one-third contained "catastrophic" themes. Moreover, when catastrophic themes were present in "ordinary" dreams, the victim seemed more capable of dealing with danger than in a comparable nightmare dream. However, in the nightmare dream, danger to the victim assumes the form of death, physical injury, or social affront. In this analysis some sex differences were noted but were not considered to be significant.

Although the above-mentioned content analyses add to our knowledge of nightmare phenomena, reported nightmare themes suffer from effects of long-term retrospective distortion and contamination by intervening daytime experiences. In order to counteract these extraneous influences, experimental sleep studies along the lines of Fisher et al. (1970) and Hartmann (1970), in which the nightmare dreamer is awakened and immediately interviewed following his nightmare, seem to offer a more promising means of obtaining minimally contaminated nightmare content. But here too, attempts to uncover retrospective distortion are subject to equally contaminating influences of experimenter suggestion. An excellent example of the combination of retrospective distortion and examiner contamination is provided by Gastaut and Broughton (1965):

> In questioning our subjects immediately after an attack, we have encountered such interpretations and have been able to study the development of events in inverse order. One of our patients suddenly awoke with a cry in intense pain. When more calm, he stated that he was dreaming of being "locked up in a tomb" unable to breathe. Questioned on the visual, tactile, or auditory images that allowed him to identify it as a tomb, he admitted that it might have been, for instance, a dark room. Questioned further, he ended by recognizing that he could hear or see nothing but was afraid of dying because of his inability to breathe or move [pp. 214–215].

Fisher, Byrne, and Edwards (1968) and Fisher et al. (1970) described typical content obtained from Stage 2 sleep, Stage 4 sleep, and REM nightmares, and have made some preliminary comparisons. More specifically, Fisher et al. (1968) found that "In some instances the NREM [non-rapid-eye-movement] nightmare and REM [rapid-eye-movement] dreams of the same night appear to deal with related content [p. 222]." Fisher et al. (1970) maintained that content of Stage 2 nightmare dreams is intermediate between the internally directed aggression of Stage 4 nightmares and the somewhat less threatening thematic material found in REM anxiety dreams. It is specified that in the Stage 4 dream (incubus attack) themes of "falling, being crushed, choking," etc., occur simultaneously with screams and other vocalizations exhibited by the nightmare sufferer. However, very similar content is often elicited from subjects awakening from REM nightmares, but subjective fright and accompanying physiological manifestations are diminished. Moreover, REM nightmares frequently represent continuations of nonnightmare REM dreams. Research involving controlled evaluations appears warranted in statistically differentiating thematic material obtained from the three kinds of nightmare experiences.

## PERSONALITY CORRELATES OF NIGHTMARE SUFFERERS

There are only four studies known to the author in which personality factors have been correlated with frequency of reported nightmares (Feldman & Hersen, 1967; Hersen, 1971a; Lester, 1968, 1969). And, in these, the results appear to be contradictory. Feldman and Hersen (1967) reported that a significant positive relationship was found between a 10-item death scale and frequency of nightmare dreams. It was also suggested that nightmare behavior and attitudes toward death might have their origins in early concrete experiences, as frequent nightmare subjects had histories of larger numbers of deaths of close friends and relatives when under 10 years of age. In a replication study, in which psychotic psychiatric inpatients were assessed, Hersen (1971a) also found a positive relationship between nightmare frequency and conscious concerns about death.

Lester (1968), in a replication of Feldman and Hersen's (1967) original study, using the Collett and Lester (1969) death scale and considering nightmare frequency in the context of reported dream frequency, failed to confirm their results. However, Lester (1968) noted that his results might be biased since the fear of death questionnaire used was different and that only female subjects were examined. In addition, it is pointed out that many subjects did not complete their questionnaires accurately. Further, in computing his major statistical analyses, Lester's (1968) high-frequency nightmare subjects were defined as those experiencing nightmares only on a once a month basis. Inasmuch as Feldman and Hersen (1967) operationally defined their frequent group as having nightmares on at least a once a week basis, it is clear that the two studies are not directly comparable. In a subsequent study, in which 170 female undergraduates served as subjects, the association between frequency of nightmares and fear of death was not found (Lester, 1969). Memory for nightmares also was not correlated with reported fears of dying. But, here too, methodology is not comparable to that used by Feldman and Hersen (1967).

Hersen (1971a), examining the relationship between nightmare frequency and personality characteristics for an inpatient psychiatric sample, found that frequent nightmare sufferers demonstrated greater manifest anxiety, lower ego-strength, and significantly more sleep disturbances. In addition, a low positive relationship between total score on a fear inventory and nightmare frequency was noted. From these results it was tentatively concluded that frequent nightmare subjects (inpatient psychiatric sample) evidence a heightened degree of pathology, but that generalizations to other populations were unwarranted.

With respect to the association of nightmare frequency and elevated levels of manifest anxiety found by Hersen (1971a), Jones (1949) previously had stressed the element of anxiety in his theoretical formulations. Gastaut and Broughton (1965), in characterizing their six male laboratory nightmare subjects, psychiatrically had assessed four as evidencing daytime symptoms of irritability and anxiety. Two other subjects were diagnosed as anxiety neurotics.

As to the marked relationship between self-reported sleep disturbances and frequency of self-reported nightmares, Hersen (1971a) depicted his frequent nightmare sufferer as experiencing related and unrelated sleep disturbances including somnambulism, enuresis, awakening in various disturbed physiological states, restless sleep, and general fearfulness about sleep. His findings are consistent with a recently conducted physiological investigation in which Broughton (1968) demonstrated the association of nightmares, nocturnal enuresis, somnambulism, and night terrors. These disturbances all seem to arise in Stage 4 sleep, *but not necessarily in the same subjects.*

Although some independent evidence exists for relationships between nightmare frequency and personality attributes, an inherent difficulty in utilizing self-report measures is that subjects who admit to negative attributes about themselves on one measure show a high probability of reporting problems and negative attributes on other self-report inventories. Thus, further objective studies that do not involve questionnaires and self-reports (e.g., experimental sleep studies, physiological measurements, behavioral situations) are necessary to bolster and confirm the obtained relationships on self-report devices. This is imperative if a viable nomological network for the nightmare personality is to be established.

## PSYCHOPHYSIOLOGY OF NIGHTMARES

Considerable data about sleep and dreaming have been amassed in the last decade and one-half since Aserinsky,[2] Aserinsky and Kleitman (1953), and Dement and Kleitman (1957) conducted their pioneering efforts relating rapid eye movement to sleep and dreaming. However, only more recently have physiological researchers turned their attention to the intriguing processes underlying the production of nightmares (Broughton, 1968, 1970; Fisher, Byrne, Edwards, & Kahn, 1969; Fisher et al., 1968; Gastaut & Broughton, 1963, 1965; Hartmann, 1970).

Broughton (1968, 1970) and Gastaut and Broughton (1965) demonstrated that adult nightmares (incubus attacks) typically occur during the arousal from slow-wave sleep, Stage 4, either in the first or second sleep cycle. Polygraphic studies conducted by Gastaut and Broughton (1965) revealed the occurrence of the following sequence of physiological events from the inception of the nightmare dream: "a K complex or burst of delta activity, succeeded in turn by an ample alpha rhythm, several ocular movements, blinking, very significant tachycardia, hyperpnea, polypnea, and movement followed by intense global muscular contraction [p. 216]." Fisher et al. (1970), in describing the subjective concomitants of the nightmare reaction, referred to a "sudden cataclysmic breakthrough of panic with intense vocalization or bloodcurdling screams [p. 184]." Broughton (1968), based on his psychophysiological studies, characterized the adult nightmare sufferer as being subject to hyperreactive heart rates during all arousal sleep stages throughout the night. In addition, it is indicated that nightmare subjects are prone to exhibiting relative tachycardia during slow-wave sleep.

It is of particular interest that incubus attacks do not in any way effect distribution of REM sleep throughout the night. On this basis, Broughton (1970) rejected the notion that physiological or psychological components of REM sleep are responsible for incubus attacks. Indeed, he contended that the incubus attack, a disorder of arousal, results from unresolved conflicts which then lead to concomitant physiological changes predisposing an episode during slow-wave-sleep arousal.

Fisher et al. (1968, 1969, 1970) have confirmed and extended Broughton's (1968, 1970) findings, but also presented contradictory evidence. Fisher et al. (1968) examined 21 experimental sleep subjects, and reported that 4 produced a total of 34 nightmares. Polygraphic readings confirmed Broughton's (1968) findings that NREM nightmares originate in Stage 4 slow-wave sleep. It was found that 73% of the nightmares took place in the initial NREM period with concomitants of autonomic discharge and screams. Although NREM sleep is depicted as being disturbed, before the beginning of an NREM nightmare episode, heart and breathing rates are within normal limits. With respect to REM nightmares or anxiety dreams, it was noted that they occur less frequently than NREM nightmares. Thus, contrary to the original findings reported by Gastaut and Broughton (1965) and subsequently Broughton (1968, 1970), there is evidence that some nightmares have their inception in REM sleep (Fisher et al., 1968, 1969, 1970; Hartmann, 1970; Jouvet, 1962; Oswald, 1968).

---

[2] E. Aserinsky. Ocular motility during sleep and its application to the study of "rest-activity" cycles and dreaming. Unpublished doctoral dissertation, University of Chicago, 1953.

In a further report, in which 37 nightmare subjects were observed for 150 nights, Fisher et al. (1969) indicated that severity of NREM nightmares and their physiological concomitants is related to depth of Stage 4 sleep directly prior to their inception. By contrast, REM nightmares (anxiety dreams) did not have unusual accompanying cardiorespiratory increases nor were vocalizations a part of the sequence. In comparing 22 spontaneous awakenings from REM sleep in 11 subjects with nonwaking REM, it was found that in 18 of the awakenings heart and respiratory rates did not differ. Fisher et al. (1969) therefore concluded: "The REM anxiety dream deals with controlled anxiety and is desomatized, and the stage 4 nightmare with uncontrolled anxiety and resomatization, e.g., intense autonomic discharge [p. 252]."

Fisher et al. (1970) were not able to confirm Broughton's (1968) suggestion that nightmare subjects demonstrate heightened physiological responsivity throughout their sleep. As indicated in their earlier study (Fisher et al., 1969), a decrease in physiological activation was observed prior to the occurrence of a Stage 4 nightmare (incubus attack). For example, the subjects' heart rates ranged from 44 to 54 per minute during Stage 4 sleep as opposed to the norm of 60 per minute. Also contrary to Broughton's (1968) initial report that there was amnesia for content of Stage 4 nightmares, Fisher et al. (1970) were able to obtain reports of content in 83% of Stage 4 nightmares observed.

In addition to Stage 4 nightmares and REM nightmares, Fisher et al. (1970) reported that some nightmares occur in Stage 2 sleep. However, there are few arousals in Stage 2 nightmares, but when they do take place it is in a subject who also experiences Stage 4 nightmares. In Stage 2 nightmares anxiety is moderate while cardiorespiratory response is greater for REM nightmares. Primary process mechanisms are thought to be used in all three types of nightmare dreams. With respect to the latter assertion, it is not at all clear to the present reviewer why the authors found it necessary to interject psychoanalytic constructions in a report that is mainly empirical in nature and psychophysiological in its orientation.

Hartmann (1970) categorized nightmares as being either the paralyzing kind (incubus) occurring early in sleep during arousal from slow-wave sleep or frightening REM (D-state nightmare) occurring later during the night. Commenting on the infrequent occurrence of nightmares observed in his sleep laboratory, Hartmann stated that in more than 500 nights' observation for normal adult subjects who were not being medicated, no incubus attacks were reported. Only one D-type nightmare was reported by a young woman expecting her menstrual period. In a number of other studies subjects were awakened from 5 to 10 minutes succeeding onset of D-period, but in over 400 awakenings no reports of nightmares were made. However, a bias may exist, as all subjects were normal, young, psychiatrically, and psychologically screened undergraduates. Most subjects were male.

In studies evaluating pathological conditions and effects of pharmacological agents, Hartmann (1970) reported that in 500 nights of sleep observation, 11 identifiable nightmares resulting in spontaneous awakenings were recorded. Of the 11, only 1 appeared to be an incubus attack, but the remaining 10 were REM nightmares. It is pointed out that the REM nightmare usually is manifested after a long and active D-period when D pressure is presumed to be at its maximum.[3] Consequently, it is hypothesized that long D-periods and increased D-pressure are related to the production of nightmares. Clinical evidence is presented in the following section that increased D-pressure by drugs and by other agents interfering with homeostasis is frequently associated with the instigation of nightmare episodes.

---

[3] Hartmann used interchangeably the terms D-state nightmare and REM nightmare. D-pressure relates to the physiological need of REM-stage sleep and can be quantified in terms of percentage of REM sleep and mean eye movement density per page.

### EFFECTS OF HOMEOSTATIC IMBALANCES

Anecdotal accounts and clinical reports suggest that a variety of homeostatic imbalances lead to increased production of nightmares. For example, Grob, Harvey, Langworthy, and Lilienthal (1947) have shown that daily intramuscular administration of the potent anticholinesterase compound di-isopropyl flurophosphate resulted in development of several symptoms of the nervous system including nightmares, headaches, tremor, and insomnia. Although these symptoms were not diminished by subsequent administration of neostigmine, atropine and barbituric acid derivatives proved to be effective in reducing their intensity.

Gross and Goodenough (1968), in evaluating sleep disturbances in the acute alcoholic psychoses, found that in the prodomal phase (onset of physiological addiction to onset of hallucinations) decreased amounts of sleep are associated with increasing frequency of nightmare reports. Greenberg and Pearlman (1967) reported that one of their alcohol study subjects exhibited a severe nightmare on his second night after withdrawing from alcohol. Prior to his nightmare, polygraphic readings indicated 100% Stage 1 REM sleep. In this connection, it should be underscored that Hartmann (1970) linked manifestation of REM nightmares to REM or D-deprivation. Evidence was also presented that D-pressure is particularly high during withdrawal stages from addiction to alcohol, barbiturates, and amphetamines. It is not surprising, then, that nightmares are frequently reported during these withdrawal periods (Oswald, 1968; Oswald & Priest, 1965).

Hartmann (1970) noted that nightmares from recovery stages subsequent to febrile illness are also frequently reported. This latter point definitely lends credence to the argument of an association between nightmares and D-deprivation, particularly as febrile illnesses have been demonstrated to reduce D-time in man, thus creating increased D-pressure (Karacan, Wolff, Williams, Hursch, & Webb, 1968). In further support of his contention, Hartmann (1970) stated:

> The fact that children, who appear to have an unusually great need for D-time, typically report far more nightmares than adults also supports the notion that nightmares occur in situations characterized by high D-pressure [p. 195].

In addition, Hartmann (1970) noted that one of his young, normal, female, experimental sleep subjects, expecting onset of her menstrual period the next day, reported being awakened by a nightmare following a 25-minute D-period. This is of some relevance as Hartmann (1966) previously had shown evidence of increased D-pressure late in the menstrual cycle.

Although clinical and anecdotal data presented are interesting, firm conclusions regarding D-pressure and production of nightmares cannot be reached at this time in the absence of empirical data. It appears that Hartmann's (1970) hypothesis could be evaluated in the experimental sleep laboratory through administration of pharmacological agents known to alter D-time and D-pressure. Frequency of resulting nightmares could then be examined with a comparable placebo-control group, using a double-blind experimental design. However, ethical considerations would have to be seriously evaluated before embarking on this mode of experimentation.

### PAVOR NOCTURNUS

As was the case with the etiology of nightmares, speculation abounds with regard to causal mechanisms of pavor nocturnus, or night terrors of children (Bakwin & Bakwin, 1953; Fenichel, 1945; Pearson, 1949). However, with a few exceptions, little of an empirical or clinical experimental nature is to be found in the literature (Broughton, 1968; Dement, 1966; Gastaut & Broughton, 1963, 1965).

Although the terms nightmare and pavor nocturnus (night terrors) are used interchangeably in describing the disturbing dreams of preadolescent children, the two types of episodes can be differentiated on the basis of specific behaviors exhibited. Kanner (1957) outlined these differences and pointed out that in pavor nocturnus, the episode occurs in a somnolent twilight state, the child abruptly sits up in bed, screams inarticulately, runs and thrashes about, is unable to present an account of his distress, is unable to be calmed by his parents, projects objects and persons in the room into the dream, perspires while dreaming, shows evidence of hallucinations, but sleeps tranquilly immediately following the episode, whose duration has been known to require as long as 15 to 20 minutes. Complete amnesia for content of the dream and occurrence of the episode is usually reported.

Kanner (1957) maintained that manifestation of night terrors is strongly suggestive that "something is wrong with the child and that it is his personality that is in need of investigation and adjustment [pp. 499–500]." Bakwin and Bakwin (1957) related night terrors to unsatisfactory home environments and "emotional strain." Fenichel (1945) stated that night terrors indicate failure of ego measures and reflect the appearance of the repressed. Extrapolating only from his clinical experience, Pearson (1949) contended that without treatment, regressive changes and sexual maladjustment will be manifested in the patient's adult personality. To the contrary, Gastaut and Broughton (1965), based on clinical studies conducted in their sleep laboratory, noted:

> daytime anxiety is as rare in children presenting night terrors as it is common in adults with nightmares. Furthermore, nothing in a given subject appears to relate childhood night terrors to possible future adult nightmares [p. 213].

Indeed, Gastaut and Broughton described their seven night-terror subjects as being otherwise "psychiatrically normal."

Discrepancies in the foregoing positions could be evaluated through comparative longitudinal studies of night-terror subjects with those subjects manifesting the ordinary nightmare dream and nonnightmare subjects. Further, objective behavioral and personality evaluations of night-terror sufferers should be conducted in order to determine whether night terrors are isolated disturbances or are more representative of generalized psychopathology, as is suggested by Kanner (1957).

Broughton (1968), Gastaut and Broughton (1965), and Dement (1966) observed in their sleep laboratories that pavor nocturnus is a sleep disorder that is associated with arousal from Stage 4 sleep. Although the number of subjects studied in the sleep laboratory is generally small, Gastaut and Broughton (1965) presented conclusive evidence that, if seven children who presented seven pavor nocturnus episodes during nine polygraphic recordings, in all subjects attacks took place

> in phase IV sleep and were associated with an intense awakening reaction leading to Phase 1$^{a3}$ patterns and massive motor and vegetative components identical to those of the nightmare [p. 218].

In a separate report, Gastaut and Broughton (1963) described the violent body movement (opistothonus) seen during pavor nocturnus episodes of two experimental sleep subjects. Total amnesia for these episodes was reported.

Broughton (1968) suggested that further experimentation is needed to determine the physiological and psychological antecedents predisposing subjects to exhibit an attack of pavor nocturnus, one of the disorders of arousal. Again, the present reviewer argues that *controlled* experimentation, in which larger subject pools are used, is necessary to identify the critical triggering mechanisms for these nocturnal paroxysms. But here too, ethical considerations should be carefully considered, especially since young children would serve as subjects.

## PSYCHOGENIC THEORIES

Cason (1935) referred to the nightmare as representing "the most important type of dream because it is the most violent psychological activity commonly occurring in sleep, and it has a greater influence on the individual than any other type of dream [p. 1]." However, in spite of the nightmare dream's marked relevance to a unified theory of dreams, Freud (1950, originally published: 1900) scarcely alluded to its existence. If then, as Freud (1920) claimed, "Dreams are the removal of sleep-disturbing stimuli by way of hallucinated satisfaction [p. 110]," it is clear that the nightmare dream involves a deviation in that it does not function as the proverbial guardian of sleep. Apparently, as has been previously articulated by Feldman and Hersen (1967), "Something has gone wrong in the work of the dream which normally, according to Freud, represents an attempt at wish fulfillment designed to keep the dreamer asleep [p. 421]." In a more recently formulated theoretical position, Breger (1967), examining dreams within the context of REM research, primary process influences, memory systems analysis, and potentiating emotional feedback systems, also omitted reference to nightmare dreams.

With the exception of Fenichel (1945) and Jones (1949), most *traditional* psychoanalysts have not addressed themselves to challenges posed by the nightmare dream to Freud's theory of dreams. Indeed, Cason (1935) attributed the foregoing omissions to the exercise of "wish fulfillment" and selective perception on the part of the psychoanalysts.

Jones (1949) argued that the nightmare is a particular form of angst or anxiety attack that "is always an expression of intense, mental conflict centering about some form of 'repressed' sexual desire [p. 44]." Fenichel (1945) pointed out that evocation of the nightmare dream results as a failure of counter-cathectic mechanisms, thus permitting repressed material to emerge. Nightmares, then, are viewed as representing one of the indicators of generalized psychopathology or neurosis. However, in direct contradiction to this thesis, nightmare subjects successfully treated via behavior therapy procedures appeared to be relatively free from other neurotic tendencies (Geer & Silverman, 1967; Silverman & Geer, 1968).

## Systematic Desensitization Treatment of Nightmares

Although few reports concerning specific therapeutic procedures to eliminate nightmares are found in the psychoanalytic literature, a number of recent case studies have appeared in which variants of Wolpe's (1958) systematic desensitization therapy have been ingeniously applied in treating pathological nightmare behavior (Cautela, 1968; Geer & Silverman, 1967; Silverman & Geer, 1968). Geer and Silverman (1967) treated a 22-year-old, married, male undergraduate student who had suffered from a recurrent nightmare (3 to 5 times a week) for the last 15 years. Aside from the nightmare there was little in his background indicating emotional maladjustment. He was a happily married veteran who had never sought any form of psychotherapeutic treatment and the Minnesota Multiphasic Personality Inventory administered resulted in $T$ scores all below 70.

In the first five therapy sessions the client was taught deep muscle relaxation à la Jacobson (1938) and was given instructions in visualization. Traditional desensitization procedures were begun, but after two sessions considerable anxiety was still elicited by imaginal presentation of dream sequences. Therefore, an adjunct of self-instructions ("It's only a dream") was used during visualization while the client attempted to remain relaxed. This combination proved to be successful in counteracting anxiety, and in a total of 13 sessions the nightmare was extinguished and no recurrence of the dream was reported. A 6-month follow-up likewise revealed no recidivism of symptoms or evidence of symptom substitution. In attempting to account for the treatment's success, Geer and Silverman speculated that self-instructions enabled the client to discriminate that occurrence of the dream was not related to reality, thus obviating necessity of the fear reaction.

The same authors (Silverman & Geer, 1968) described a second case involving a 19-year-old female undergraduate whose recurrent nightmare dream of falling off bridges had manifested itself twice weekly for approximately 4 years. Inasmuch as the patient had a related phobia, in which she experienced general fear of being "high up from the ground without maximal enclosure and support," it was decided to treat the fear of heights phobia while observing the course of the nightmare. Using systematic desensitization therapy with a six-item fear hierarchy, a total of six sessions was required to eliminate both the phobia and related nightmare. A subsequent check 1 and 6 months following brief treatment revealed no evidence of symptom return or symptom substitution.

Cautela (1968) also reported the successful treatment of recurrent nightmare dreams for three patients via short-term systematic desensitization therapy.

Although the case histories reviewed tentatively suggest that extinction of nightmares is amenable to aspects of systematic desensitization therapy, it is not clear which components of the treatment are the responsible agents of change. Consequently, Baron and Cautela[4] have designed an experiment in which treatment factors could be isolated: "a desensitization group; a relaxation only group; an attention placebo group; a no-content group," However, since in their sample only 2% admit to experiencing frequent nightmares (at least one per week), enough subjects have not yet been gathered to complete the experiment.

### Discussion and Conclusions

In many respects, the studies reported are of dubious reliability and validity, conclusions are made on the basis of small samples, many are clinical investigations lacking appropriate controls, and most are in need of replication with a variety of populations tapping different age levels and developmental stages. The type of surveys conducted by Cason (1935), Feldman and Hersen (1967), and Hersen (1971a), although adding information about the nightmare sufferer, are beset with numerous difficulties; namely, how reliable and how valid are self-reports and self-descriptions of nightmare experiences? Discrepancies between self-reported frequency of nightmares and observable nightmare episodes in the experimental sleep laboratory are large. Hartmann (1970), for example, recorded very few nightmare episodes in his laboratory. Observations made by Zung (1970) also suggested that subjects' reports of nighttime and sleep-related experiences are poorly correlated with laboratory-observed frequency rates. Of course, it is possible that being observed in an experimental sleep laboratory has a suppressing effect on manifestations of nighttime paroxysms. However, all of the above-mentioned factors should be carefully evaluated when making conclusions about and generalizing from the obtained data.

The question of reliability of self-reports must also be examined in the larger context of "acquiescence," "yea-saying," and "nay-saying" tendencies. As was mentioned in a prior section, subjects are known to respond in specified ways independent of content when filling out questionnaires (Ford, 1964). Thus, data adduced by the self-inventory technique must be bolstered by independent reports of significant others, observations in sleep laboratories, and physiological recordings, particularly as retrospective distortion is also a potent factor in the recounting of nighttime behaviors.

The studies concerning nightmare themes are interesting, but are primarily at the descriptive level. Efforts should therefore be made to categorize thematic material gathered from nightmare reports, occurring in Stage 2 sleeep, Stage 4 sleep, and the REM stage, in order to permit quantitative comparisons.

Perhaps, the most illuminating studies are those in which the psychophysiological concomitants of nightmares have been assessed (Broughton, 1968, 1970; Fisher et al., 1968, 1969, 1970; Gastaut & Broughton, 1963, 1965; Hartmann, 1970; Jouvet, 1962; Oswald, 1968). Unfortunately, these investigations are correlational and suffer from inadequate numbers of subjects and total lack of controls. Further, as subject populations are somewhat restricted, it would appear that comparative and longitudinal studies are needed (*a*) to establish the reliability and validity of existing findings and (*b*) to determine developmental aspects of the underlying psychophysiological dimensions. As was noted in the sections on the psychophysiology of nightmares and on pavor nocturnus, research in which independent variables are manipulated through administrations of pharmacological agents, altering sleep cycles, and eliciting nightmare episodes would be invaluable. However, here the problem involving ethical considerations might provide a barrier to controlled investigations.

Finally, the work of Cautela (1968), Geer and Silverman (1967), and Silverman and Geer (1968), in which the experimental use of behavior therapy is applied to extinguish manifestation of nightmare dreams, offers promise in alleviating the nightmare sufferer of his distress. However, thus far experimental data are lacking in support of obtained clinical successes.

---

[4] M. G. Baron & J. R. Cautela. The modification of nightmares: An outcome study. Unpublished study, Boston College, 1969.

Although the nightmare dream has occupied the speculative attention of writers, philosophers, playwrights, medical doctors, and scientists since antiquity, understanding of its nature, cause, and course of progress is relatively minimal. This appears to be largely the result of "armchair" psychology, which often leads to production of untestable theoretical assumptions.

## REFERENCES

Aserinsky, E., & Kleitman, N. Regular occurring periods of eye-motility and concomitant phenomena during sleep. *Science,* 1953, **118**, 273–274.

Bakwin, H., & Bakwin, R. M. *Clinical management of behavior disorders in children.* Philadelphia: W. B. Saunders, 1953.

Bellamy, R. The analysis of a nightmare. *Journal of Abnormal Psychology* 1915, **10**, 11–18.

Blanchard, P. A study of subject matter and motivation of children's dreams. *Journal of Abnormal and Social Psychology,* 1926, **21**, 24–37.

Breger, L. Function of dreams. *Journal of Abnormal Psychology,* 1967, 72 (5, Whole No. 64).

Broughton, R. J. Sleep disorders: Disorders of arousal? *Science,* 1968, **159**, 1070–1078.

Broughton, R. J. The incubus attack. In E. Hartmann (Ed.), *Sleep and dreaming.* Boston: Little, Brown, 1970.

Cason, H. The nightmare dream. *Psychological Monographs,* 1935, **46** (5, Whole No. 209).

Cautela, J. R. Behavior therapy and the need for behavioral assessment. *Psychotherapy,* 1968, **5**, 175–179.

Collett, L. J., & Lester, D. The fear of death and the fear of dying. *Journal of Psychology,* 1969, **72**, 179–181.

Dement, W. C. Psychophysiology of sleep and dreams. In S. Arieti (Ed.), *American handbook of psychiatry. Vol. III.* New York: Basic Books, 1966.

Dement, W. C., & Kleitman, N. The relation of eye-movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology,* 1957, **53**, 339–346.

Feldman, M. J., & Hersen, M. Attitudes toward death in nightmare subjects. *Journal of Abnormal Psychology,* 1967, **72**, 421–425.

Feldman, M. J., & Hyman, E. Content analysis of nightmare reports. *Psychophysiology,* 1968, **5**, 221.

Fenichel, O. *The psychoanalytic theory of neurosis.* New York: W. W. Norton, 1945.

Fisher, C., Byrne, J. V., & Edwards, A. NREM and REM nightmares. *Psychophysiology,* 1968, **5**, 221–222.

Fisher, C., Byrne, J., Edwards, A., & Kahn, E. REM and NREM nightmares. *Psychophysiology,* 1969, **6**, 252.

Fisher, C., Byrne, J. V., Edwards, A., & Kahn, E. REM and NREM nightmares. In E. Hartmann (Ed.), *Sleep and dreaming.* Boston: Little, Brown, 1970.

Ford, L. H., Jr. A forced-choice acquiescence-free, social desirability (defensiveness) scale. *Journal of Consulting Psychology,* 1964, **28**, 475.

Freud, S. *A general introduction to psychoanalysis.* New York: Boni & Liveright, 1920.

Freud, S. *The interpretation of dreams.* New York: Modern Library, 1950. [*Die Traumdeutung.* Leipzig: F. Deuticke, 1900.]

Gastaut, H., & Broughton, R. Paroxysmal psychological events and certain phases of sleep. *Perceptual and Motor Skills,* 1963, **17**, 362.

Gastaut, H., & Broughton, R. A clinical and polygraphic study of episodic phenomena during sleep. In J. Vostis (Ed.), *Recent advances in biological psychiatry.* Vol. 8. New York: Plenum Press, 1965.

Geer, J. H. The development of a scale to measure fear. *Behavior Research and Therapy,* 1965, **3**, 45–53.

Geer, J. H., & Silverman, I. Treatment of a recurrent nightmare by behavior modification procedures. *Journal of Abnormal Psychology,* 1967, **72**, 188–190.

Greenberg, R., & Pearlman, C. Delirium tremens and dreaming. *American Journal of Psychiatry,* 1967, **124**, 37–46.

Grob, H. A., Harvey, O. R., Langworthy, O., & Lilienthal, J. The administration of di-isopropyl fluorophosphate (DFP) to man. *Bulletin of the Hopkins Hospital,* 1947, **81**, 257–266.

Gross, M. M., & Goodenough, D. R. Sleep disturbances in the acute alcoholic psychoses. In, *Psychiatric Research Report 24.* Washington, D. C.: American Psychiatric Association, 1968.

Hadfield, J. A. *Dreams and nightmares.* London: Penguin Books, 1954.

Hartmann, E. Dreaming sleep (the D-state) and the menstrual cycle. *Journal of Nervous and Mental Disease,* 1966, **143**, 406–416.

Hartmann, E. A note on the nightmare. In E. Hartmann (Ed.), *Sleep and dreaming,* Boston: Little, Brown, 1970.

Hersen, M. Personality characteristics of nightmare sufferers. *Journal of Nervous and Mental Disease,* 1971, **153**, 27–31. (a)

Hersen, M. Fear scale norms for an in-patient population. *Journal of Clinical Psychology,* 1971, **27**, 375–378. (b).

Jacobson, E. *Progressive relaxation.* Chicago: University of Chicago Press, 1938.

Jones, E. *On the nightmare.* London: Hogarth Press, 1949.

Jouvet, M. Recherches sur les structures nerveuses et les mechanisms responsibles des differentes phase du sommeil. *Archives Italiennes de Biologie,* 1962, **100**, 125–206.

KANNER, L. *Child psychiatry*. Springfield, Ill.: Charles C Thomas, 1957.

KARACAN, I., WOLFF, L. M., WILLIAMS, R. L., HURSCH, C. J., & WEBB, W. F. The effects of fever on sleep and dream patterns. *Psychosomatics*, 1968, **9**, 331–339.

LESTER, D. The fear of death of those who have nightmares. *Journal of Psychology*, 1968, **69**, 245–247.

LESTER, D. Fear of death and nightmare experiences. *Psychological Reports*, 1969, **25**, 437–438.

OSWALD, I. Drugs and sleep. *Pharmacological Review*, 1968, **20**, 273–303.

OSWALD, I., & PRIEST, R. G. Five weeks to escape the sleeping-pill habit. *British Medical Journal*, 1965, **2**, 1093–1095.

PEARSON, G. H. J. *Emotional disorders of children: A case book of child psychiatry*. New York: W. W. Norton, 1949.

SCHNECK, J. M. Henry Fuseli, nightmare and sleep paralysis. *The Journal of The American Medical Association*, 1969, **207**, 725–726.

SILVERMAN, I., & GEER, J. M. The elimination of a recurrent nightmare by desensitization of a related phobia. *Behavior Research and Therapy*, 1968, **6**, 109–111.

STERN, A. A contribution to the localization of mental phenomena. *Schweizer Archiv Fuer Neurologie und Psychiatrie*, 1951, **67**, 405–414.

DE VERTEUIL, R. L. A psychiatric approach to the study of nightmare. *Canadian Psychiatric Association Journal*, 1962, **7**, 151–158.

WOLPE, J. *Psychotherapy by recriprocal inhibition*. Stamford: Stamford University Press, 1958.

ZUNG, W. W. K. The pharmacology of disordered sleep: A laboratory approach. In E. Hartmann (Ed.), *Sleep and dreaming*. Boston: Little, Brown, 1970.

(Received October 12, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 49–61

# AN APPROACH TO THE PSYCHOLOGY OF INSTRUCTION

R. C. ATKINSON[1] AND J. A. PAULSON

*Stanford University*

The relationship between a theory of learning and a theory of instruction is discussed. Examples are presented that illustrate how to proceed from a theoretical description of the learning process to the specification of an optimal strategy for carrying out instruction. The examples deal with fairly simple learning tasks and are admittedly of limited generality. Nevertheless, they clearly define the steps necessary for deriving and testing instructional strategies, thereby providing a set of procedures for analyzing more complex problems. The parameter-dependent optimization strategies are of particular importance because they take into account individual differences among learners as well as differences in difficulty among curriculum units. Experimental evaluations indicate that the parameter-dependent strategies lead to major gains in learning, when compared with strategies that do not take individual differences into account.

The task of relating the methods and findings of research in the behavioral sciences to the problems of education is a continuing concern of both psychologists and educators. A few years ago, when our faith in the ability of money and science to cure social ills was at its peak, an educational researcher could content himself with trying to answer the same questions that were being studied by his psychologist colleagues. The essential difference was that his studies referred explicitly to educational settings, whereas those undertaken by psychologists strived for greater theoretical generality. There was implicit confidence that as the body of behavioral research grew, applications to education would occur in the natural course of events. When these applications failed to materialize, confidence was shaken. Clearly, something essential was missing from educational research.

A number of factors contributed to the feeling that something was wrong with business as usual. Substantial curriculum changes, initiated on a national scale after the Soviet's launching of Sputnik, had to be carried out with only minimal guidance from behavioral scientists. Developers of programmed learning and computer-assisted instruction faced similar problems. Although the literature in learning theory was perhaps more relevant to their concerns, the questions it treated were still not the critical ones from the viewpoint of instruction. This situation would not have been surprising had the study of learning been in its infancy. But far from that, the psychology of learning had a long and impressive history. An extensive body of experimental literature existed, and many simple learning processes were being described with surprising precision using mathematical models. Whatever was wrong, it did not seem to be a lack of scientific sophistication.

These issues were on the minds of those who contributed to the 1964 *Yearbook of the National Society for the Study of Education*, edited by Hilgard (1964). In that book Bruner (1964) summarized the feelings of many of the contributors when he called for a theory of instruction, which he sharply distinguished from a theory of learning. He emphasized that where the latter is essentially descriptive, the former should be prescriptive, setting forth rules specifying the most effective ways of achieving knowledge or mastering skills. This distinction served to highlight the difference in the goals of experiments designed to advance the two kinds of theory. In many instances variations in instructional procedures affect several psychological variables simultaneously. Experiments that are appropriate for comparing methods of instruction may be virtually impossible to interpret in terms of learning theory because of this confounding of variables. The importance of developing a theory of instruction justifies experimental programs designed to explore alternative instructional procedures, even if the resulting experiments are difficult to place in a learning-theoretic framework.

---

[1] An earlier version of this paper was presented by the first author at a conference on "The Use of Computers in Education" organized by the Japanese Ministry of Education in collaboration with the Organization for Economic Cooperation and Development in Tokyo, July 1970.

Requests for reprints should be sent to R. C. Atkinson, Department of Psychology, Stanford University, Stanford, California 94305.

The task of going from a description of the learning process to a prescription for optimizing learning must be clearly distinguished from the task of finding the appropriate theoretical description in the first place. However, there is a danger that preoccupation with finding prescriptions for instruction may cause us to overlook the critical interplay between the two enterprises. Developments in control theory (Bellman, 1961) and statistical decision theory (Raiffa & Schlaiffer, 1968) provide potentially powerful methods for discovering optimal decision-making strategies in a wide variety of contexts. In order to use these tools it is necessary to have a reasonable model of the process to be optimized. As noted earlier, some learning processes can already be described with the required degree of accuracy. This article examines an approach to the psychology of instruction which is appropriate when the learning is governed by such a process.

## STEPS IN THE DEVELOPMENT OF OPTIMAL INSTRUCTIONAL STRATEGIES

The development of optimal strategies can be broken down into a number of tasks that involve both descriptive and normative analyses. One task requires that the instructional problem be stated in a form amenable to a decision-theoretic analysis. While the detailed formulations of decision problems vary widely from field to field, the same formal elements can be found in most of them. It will be a useful starting point to identify these elements in the context of an instructional situation.

The formal elements of a decision problem that must be specified are:

1. The possible states of nature;
2. The actions that the decision maker can take to transform the state of nature;
3. The transformation of the state of nature that results from each action;
4. The cost of each action;
5. The return resulting from each state of nature.

Statistical aspects occur in a decision problem when uncertainty is associated with one or more of these elements. For example, the state of nature may be imperfectly observable or the transformation of the state of nature which a given action will cause may not be completely predictable.

In the context of the psychology of instruction, most of these elements divide naturally into two groups, those having to do with the description of the underlying learning process and those specifying the cost-benefit dimensions of the problem. The one element that does not fit is the specification of the set of actions from which the decision maker must make his choice. The nature of this element can be indicated by an example.

Suppose one wants to design a supplemental program of exercises for an initial reading program. Most reasonable programs of initial reading instruction include both training in sight-word identification and training in phonics. Let us assume that on the basis of experimentation two useful exercise formats have been developed, one for training on sight words, the other for phonics. Given these formats, there are many ways to design an overall program. A variety of optimization problems can be generated by fixing some features of the design and leaving the others to be determined in a theoretically optimal manner. For example, it may be desirable to determine how the time available for instruction should be divided between phonics and sight-word recognition, with all other features of the design fixed. A more complicated question would be to determine the optimal ordering of the two types of exercises in addition to the optimal allocation of time. It would be easy to continue generating different optimization problems in this manner. The point is that varying the set of actions from which the decision maker is free to choose changes the decision problem, even though the other elements remain the same.

For the decision problems that arise in instruction it is usually natural to identify the states of nature with learning states of the student. Specifying the transformation of the states of nature caused by the actions of the decision maker is tantamount to constructing a model of learning for the situation under consideration.

The role of costs and returns is more formal than substantive for the class of decision problems considered in this article. The specification of costs and returns in instructional situations tends to be straightforward when examined on a short-time basis, but virtually intractable over the long term. In the short term, one can assign costs and returns for the mastery of, say, certain basic reading skills, but sophisticated determinations for the long-term value of these skills to the individual and society are difficult to make. There is an important role for detailed economic analysis of the long-term impact of education, but such studies deal with issues at a more global level than we require. In this article analysis is limited to those costs and returns directly related to the specific instructional task being considered.

After a problem has been formulated in a way amenable to decision-theoretic analysis, the next step is to derive the optimal strategy for the learning model which best describes the situation. If more than one learning model seems reasonable a priori, then competing candidates for the optimal strategy can be deduced. When these steps have been accomplished, an experiment can be designed to determine which strategy is best.

There are several possible directions in which to proceed after the initial comparison of strategies, depending on the results of the experiment. If none of the supposedly optimal strategies produces satisfactory results, then further experimental analysis of the assumptions of the underlying learning models is indicated. New issues may arise even if one of the procedures is successful. In one case that we discuss, the successful strategy produced an usually high error rate during learning, which is contrary to a widely accepted principle of programmed instruction. When anomalies such as this occur, they suggest new lines of experimental inquiry, and often require a reformulation of the axioms of the learning model. The learning model may have provided an excellent account of data for a range of experimental conditions but can prove totally inadequate in an optimization condition where special features of the procedure magnify inaccuracies of the model that had previously gone undetected.

## AN OPTIMIZATION PROBLEM THAT ARISES IN COMPUTER-ASSISTED INSTRUCTION

One application of computer-assisted instruction that has proved to be very effective in the primary grades involves a regular program of practice and review specifically designed to complement the efforts of the classroom teacher (Atkinson, 1969). Some of the curriculum materials in such programs take the form of lists of instructional units or items. The objective of the computer-assisted instruction programs is to teach students the correct response to each item in a given list. Typically, a sublist of items is presented each day in one or more fixed exercise formats. The optimization problem that arises concerns the selection of items for presentation on a given day.

The Stanford Reading Project is an example of such a program in initial reading instruction (Atkinson & Fletcher, 1972). The vocabularies of several of the commonly used basal readers were compiled into one dictionary, and a variety of exercises using these words were developed to teach reading skills. These exercises were designed principally to strengthen the student's decoding skills, with special emphasis on letter identification, sight-word recognition, phonics, spelling patterns, and word comprehension. The details of the teaching procedure vary from one exercise to another, but most include a sequence in which a curriculum item is presented, eliciting a response from the student, followed by a short period for studying the correct response. For example, one exercise in sight-word recognition has the following format:

| Teletype Display | Audio Message |
|---|---|
| NUT MEN RED | Type red. |

Three words are printed on the teletype, followed by an audio presentation of one of the words. Control is then turned over to the student; if he types the correct word a reinforcing message is given, and the computer program then proceeds to the next presentation. If the student responds incorrectly or exceeds the time, the teletype prints the correct word simultaneously with its audio presentation and then moves to the next presentation. Under an early version of the program, items were presented in predetermined sublists, with an exercise continuing on a sublist until a specified criterion has been met.

Strategies can be found that will improve on the fixed order of presentation. Two studies described below are concerned with the development of such strategies. One study examines alternative presentation strategies for teaching spelling words to elementary school children, and the other examines strategies for teaching Swahili vocabulary items to college-level students. The optimization problems in both studies were essentially the same. A list of $N$ items is to be learned, and a fixed number of days, $D$, are allocated for its study. On each day a sublist of items is presented for test and study. The sublist always involves $M$ items, and each is presented only once for test followed by a study period. The total set of $N$ items is extremely large with regard to any sublist of $M$ items. Once the experimenter has specified a sublist for a given day, its order of presentation is random. After the $D$ days of study are completed, a posttest is given over all items. The parameters $N$, $D$, and $M$ are fixed, and so is the instructional format on each day. Within these constraints the problem is to maximize performance on a posttest by an appropriate selection of sublists from day to day. The strategy for selecting sublists from day to day is dynamic (or *response sensitive*, using the terminology of Groen & Atkinson, 1966) to the extent that it depends upon the student's prior history of performance.

### *Three Models of the Learning Process*

Two extremely simple learning models are considered first. Then a third model, which combines features of the first two, is described.

In the first model, the state of the learner with respect to each item is completely determined by the number of times the item has been studied. At the start of the experiment an item has some initial probability of error; each *time* the item is presented its error probability is reduced by a factor $\alpha$, which is less than one. Stated as a difference equation, the probability of an error on the $n + 1$st presentation of an item is related to its probability on the $n$th presentation as follows:

$$q_{n+1} = \alpha q_n. \qquad [1]$$

Note that the error probability for a given item depends on the number of times it has been reduced by the factor $\alpha$; that is, the number of times it has been presented. Learning is the gradual reduction in the probability of error by repeated presentations of items. This model is sometimes called the *linear model* because the equation describing change in response probability is linear.

In the second model, mastery of an item is not at all gradual. At any point in time a student is in one of two states with respect to each item: the learned state or the unlearned state. If an item in the learned state is presented, the correct response is always given; if an item is in the unlearned state, an incorrect response is given unless the student makes a correct response by guessing. When an unlearned item is presented, it may move into the learned state with probability $c$. Stated as a difference equation,

$$q_{n+1} = \begin{cases} q_n, \text{ with probability } 1 - c \\ 0, \text{ with probability } c. \end{cases} \qquad [2]$$

Once an item is learned, it remains in the learned state throughout the course of instruction. Some items are learned the first time they are presented; others may be presented several times before they are finally learned. Therefore, the list as a whole is learned gradually. But for any particular item, the transition from the unlearned to the learned state occurs on a single trial. The model is sometimes called the *all-or-none model* because of this characterization of the possible states of learning.

The third model to be considered is the *random-trial increments model* and represents a compromise between the linear and all-or-none model (Norman, 1964). For this model:

$$q_{n+1} = \begin{cases} q_n, \text{ with probability } 1 - c \\ \alpha q_n, \text{ with probability } c. \end{cases} \qquad [3]$$

If $c = 1$, the random-trial increments model reduces to the linear model; if $\alpha = 0$, it reduces to the all-or-none model. However, for $c < 1$ and $\alpha > 0$, the random-trial increments model generates predictions that are quite distinct from both the linear and the all-or-none models.

For all three models the probability of an error on the first trial is a parameter that may need to be estimated in certain situations; to emphasize this point the initial error probability is written as $q'$ henceforth. It should be noted that the all-or-none model and the random-trial increments model are *response sensitive* in that the learner's particular history of correct and incorrect responses makes a difference in predicting performance on the next presentation of an item. In contrast, the linear model is *response insensitive*; its prediction depends only on the number of prior presentations and is not improved by a knowledge of the learner's response history.

### *Cost/Benefit Structure*

At the present level of analysis, it will expedite matters if some assumptions are made to simplify the appraisal of costs and benefits associated with various strategies. It is tacitly assumed that the subject matter being taught is sufficiently beneficial to justify allocating a fixed amount of time to it for instruction. Since the exercise formats and the time allocated to instruction are the same for all strategies, it is reasonable to assume that the costs of instruction are the same for all strategies as well. If the costs of instruction are equal for all strategies, then for purposes of comparison they may be ignored and attention focused on the comparative benefits of the various strategies. This is an important simplification because it affects the degree of precision necessary in the assessment of costs and benefits. If both costs and benefits are significantly variable in a problem, then it is essential that both quantities be estimated accurately. This is often difficult to do. When one of these quantities can be ignored, it suffices if the other can be assessed accurately enough to order the possible outcomes. This is usually fairly easy to accomplish. In the present problem, for example, it is reasonable to consider all the items equally beneficial. This implies that benefits depend only on the overall probability of a correct response, not on the particular items known. It turns out that this specification of cost and benefit is sufficient for the learning models to determine optimal strategies.

The above cost/benefit assumptions permit us to concentrate on the main concern of this article, the derivation of the educational implications of learning models. Also, they are approximately valid in many instructional contexts. Nevertheless, it must be recognized that in the majority of cases these assumptions will not be satisfied. For instance, the assumption that the alternative strategies cost the same to implement usually does not hold. It only holds as a first approximation in the case being considered here. In the present formulation of the problem, a fixed amount of time is allocated for study and the problem is to maximize learning, subject to this time constraint. An alternative formulation that is more appropriate in some situations fixes a minimum criterion level for learning. In this formulation, the problem is to find a strategy for achieving this criterion level of performance in the shortest time. As a rule, both costs and benefits must be weighed in the analysis, and frequently subtopics within a curriculum vary significantly in their importance. Sometimes there is a choice among several exercise formats. In certain cases, whether or not a certain topic should be taught at all is the critical question. Smallwood (1971) has treated a problem similar to the one considered in this article in a way that includes some of these factors in the structure of costs and benefits.

### *Deducing Strategies from the Learning Models*

Optimal strategies can be deduced for the linear and all-or-none models under the assumption that all items have the same learning parameters and initial error probabilities. The situation is more complicated in the case of the random-trial increments model. An approximation to the optimal strategy for the random-trial increments case will be discussed later; in this case the strategy explicitly allows for individual differences in parameter values.

For the linear model, if an item has been presented $n$ times, the probability of an error on the next presentation of the item is $\alpha^{n-1}q'$; when the item is presented, the error probability is reduced to $\alpha^n q'$. The size of the reduction is thus $\alpha^{n-1}(1-\alpha)q'$. Observe that the size of the decrement in error probability gets smaller with each presentation of the item. This observation can be used to deduce that the following procedure is optimal:

> On a given day, form the sublist of $M$ items by selecting those items that have received the fewest presentations up to that point. If more than $M$ items satisfy this criterion, then select items at random from the set satisfying the criterion.

Upon examination, this strategy is seen to be equivalent to the standard cyclic presentation procedure commonly employed in experiments on paired-associate learning. It amounts to presenting all items once, randomly reordering them, presenting them again, and repeating the procedure until the number of days allocated to instruction have been exhausted.

According to the all-or-none model, once an item has been learned there is no further reason to present it. Since all unlearned items are equally likely to be learned if presented, it is intuitively reasonable that the optimal presentation strategy selects the item least likely to be in the learned state for presentation. In order to discover a good index of the likelihood of being in the learned state, consider a student's response protocol for a single item. If the last response was incorrect, the item was certainly in the unlearned state at that time, although it may then have been learned during the study period that immediately followed. If the last response was correct, then it is more likely that the item is now in the learned state. In general, the more correct responses there are in the protocol since the last error on the item, the more likely it is that the item is in the learned state.

The preceding observations provide a heuristic justification for an algorithm which Karush and Dear (1966) have proved is in fact the optimal strategy for the all-or-none model. The optimal strategy requires that for each student a bank of counters be set up, one for each word in the list. To start, $M$ different items are presented each day until each item has been presented once and a 0 has been entered in its counter. On all subsequent days, the strategy requires that we conform to the following two rules:

> 1. Whenever an item is presented, increase its counter by 1 if the subject's response is correct, but reset it to 0 if the response is incorrect.
> 2. Present the $M$ items whose counters are lowest among all items. If more than $M$ items are eligible, then select randomly as many items as are needed to complete the sublist of size $M$ from those having the same highest counter reading, having selected all items with lower counter values.

For example, suppose six items are presented each day, and after a given day a certain student has four items whose counters are 0, four whose counters are 1, and higher values for the rest of the counters. His study list would consist of the four items whose counters are 0, and two items selected at random from the four whose counters are 1.

It has been possible to find relatively simple optimal strategies for the linear and all-or-none models. It is noteworthy that neither strategy depends on the values of the parameters of the respective models (i.e., on $\alpha$, $c$, or $q'$). Another exceptional feature of these two models is that it is possible to condense a student's response protocol to one index per item without losing any information relevant to presentation decisions. Such condensations of response protocols are referred to as *sufficient histories* (Groen & Atkinson, 1966). Roughly speaking, an index summarizing the information in a student's response protocol is a sufficient history if any additional information from the protocol would be redundant in the determination of the student's state of learning. The concept is analogous to a *sufficient statistic*. If one takes a sample of observations from a population with an underlying normal distribution and wishes to estimate the population mean, the sample mean is a sufficient statistic. Other statistics that can be calculated (such as the median, the range, and the standard deviation) cannot be used to improve on the sample mean as an estimate of the population mean, though they may be useful in assessing the precision of the estimate. In statistics, whether or not data can be summarized by a few simple sufficient statistics is determined by the nature of the underlying distribution. For educational applications, whether or not a given instructional process can be adequately monitored by a simple sufficient history is determined by the model representing the underlying learning process.

The random-trial increments model appears to be an example of a process for which the information in the subject's response protocol cannot be condensed into a simple sufficient history. It is also a model for which the optimal strategy depends on the values of the model parameters. Consequently, it is not possible to state a simple algorithm for the optimal presentation strategy for this model. Suffice it to say that there is an easily computable formula for determining which item has the best expected immediate gain, if presented. The strategy that presents this item should be a reasonable approximation to the optimal strategy (Calfee, 1970). More is said later regarding the problem of parameter estimation and some of its ramifications.

If the three models under consideration are to be ranked on the basis of their ability to account for data from laboratory experiments employing the standard presentation procedure, the order of preference is clear. The all-or-none model provides a better account of the data than the linear model, and the random-trial increments model is better than either of them (Atkinson & Crothers, 1964). This does not necessarily imply, however, that the optimization strategies derived from these models will receive the same ranking. The standard cyclic presentation procedure used in most learning experiments may mask certain deficiencies in the all-or-none or random-trial increments models which would manifest themselves when the optimal presentation strategy specified by one or the other of these models was employed.[2]

### EVALUATION OF THE ALL-OR-NONE STRATEGY

Lorton[3] compared the all-or-none strategy with the standard procedure in an experiment in computer-assisted spelling instruction with elementary school children. The former strategy is optimal if the learning process is indeed all-or-none, whereas the latter is optimal if the process is linear. The experiment was one phase of the Stanford Reading Project using computer facilities at Stanford University linked via telephone lines to student terminals in the schools.

Individual lists of 48 words were compiled in an extensive pretest program to guarantee that each student would be studying words of approximately equal difficulty, which he did not already know how to spell. A within-subjects design was used in an effort to make the comparison of strategies as sensitive as possible. Each student's individualized list of 48 words was used to form two comparable lists of 24 words, one to be taught using the all-or-none strategy and the other using the standard procedure.

Each day a student was given training on 16 words, 8 from the list for standard presentation and 8 from the list for presentation according to the all-or-none strategy. There were 24 training sessions followed by 3 days for testing all the words; approximately 2 weeks later 3 more days were spent on a delayed retention test. Using this procedure, all words in the standard presentation list received exactly one presentation in successive 3-day blocks during training. Words in the list presented according to the all-or-none algorithm received from zero to three presentations in successive 3-day blocks during training, with one presentation being the average. A flow chart of the daily routine is given in Figure 1.

The results of the experiment are summarized in Figure 2. The proportions of correct responses are plotted for successive 3-day blocks during training, followed by the first overall test and then the 2-week delayed test. Note that during training the proportion correct is always *lower* for the all-or-none procedure than for the standard procedure, but on

---

[2] This type of result was obtained by Dear, Silberman, Estavan, and Atkinson (1967). They used the all-or-none model to generate optimal presentation schedules where there were no constraints on the number of times a given item could be presented for test and study within an instructional period. Under these conditions the model generates an optimal strategy that has a high probability of repeating the same item over and over again until a string of correct responses occurs. In their experiment the all-or-none strategy proved quite unsatisfactory when compared with the standard presentation schedule. The problem was that the all-or-none model provides an accurate account of learning when the items are well spaced, but fails badly under highly massed conditions. Laboratory experiments prior to the Dear et al. (1967) study had not employed a massing procedure, and this particular deficiency of the all-or-none model had not been made apparent. The important remark here is that the analysis of instructional problems can provide important information in the development of learning models. In certain cases the set of laboratory tasks that the psychologist deals with may be such that it fails to uncover that particular condition which would cause the model to fail. By analyzing optimal learning conditions we are imposing a somewhat different test on a learning model, which may provide a more sensitive measure of its adequacy.

[3] P. Lorton, Jr. Computer-based instruction in spelling: An investigation of optimal strategies for presenting instructional material. Unpublished doctoral dissertation, Stanford University, 1972.

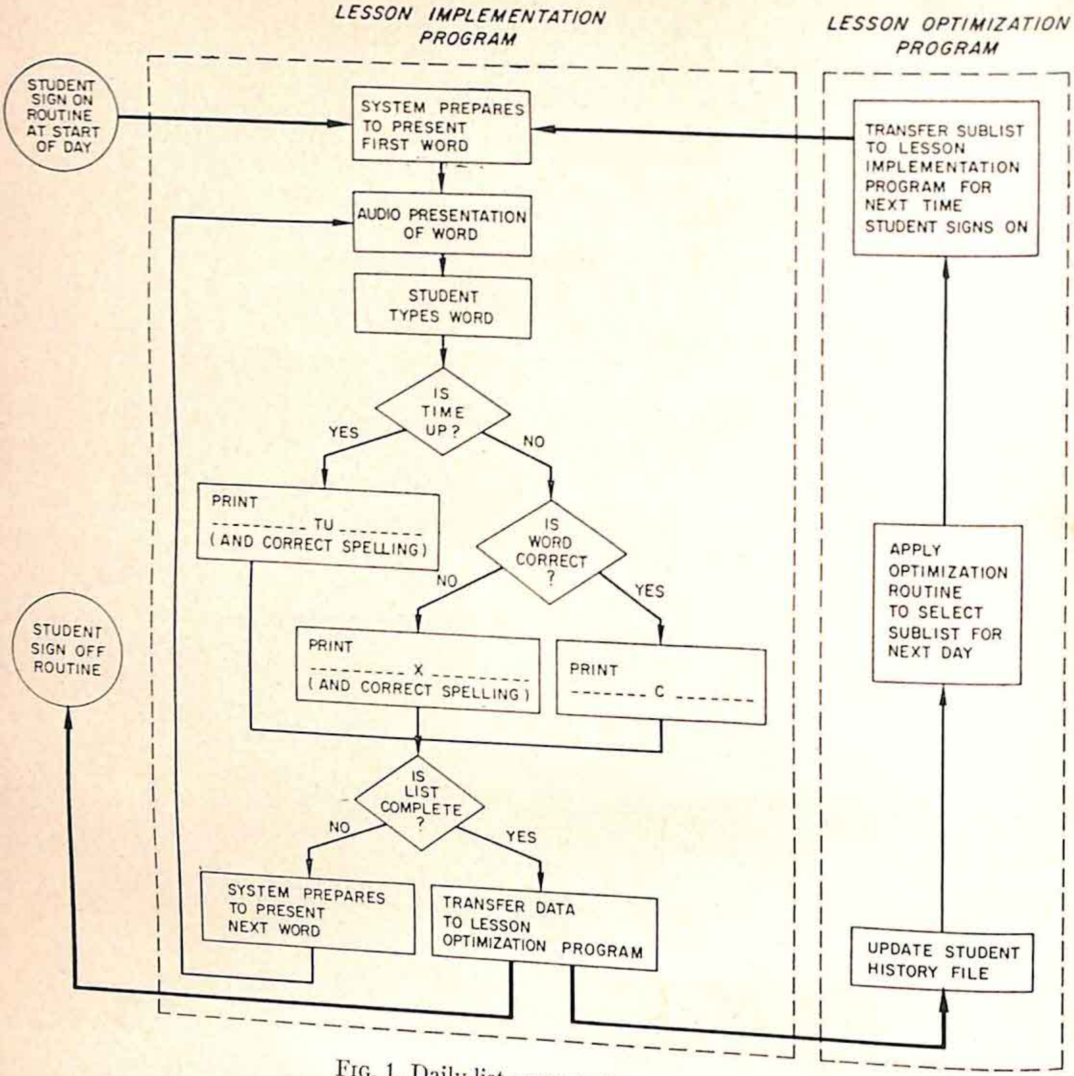

LESSON IMPLEMENTATION PROGRAM
LESSON OPTIMIZATION PROGRAM
STUDENT SIGN ON ROUTINE AT START OF DAY
SYSTEM PREPARES TO PRESENT FIRST WORD
TRANSFER SUBLIST TO LESSON IMPLEMENTATION PROGRAM FOR NEXT TIME STUDENT SIGNS ON
AUDIO PRESENTATION OF WORD
STUDENT TYPES WORD
IS TIME UP?
YES
NO
PRINT --------- TU -------- (AND CORRECT SPELLING)
IS WORD CORRECT?
NO
YES
APPLY OPTIMIZATION ROUTINE TO SELECT SUBLIST FOR NEXT DAY
STUDENT SIGN OFF ROUTINE
PRINT --------- X --------- (AND CORRECT SPELLING)
PRINT ------- C -------
IS LIST COMPLETE?
NO
YES
SYSTEM PREPARES TO PRESENT NEXT WORD
TRANSFER DATA TO LESSON OPTIMIZATION PROGRAM
UPDATE STUDENT HISTORY FILE

FIG. 1. Daily list presentation routine.

both the final test and the retention test the proportion correct is *greater* for the all-or-none strategy. Analysis of variance tests verified that these results are statistically significant. The advantage of approximately 10 percentage points on the posttests for the all-or-none procedure is of practical significance as well.

The observed pattern of results is exactly what would be predicted if the all-or-none model does indeed describe the learning process. As was shown earlier, final test performance should be better when the all-or-none optimization strategy is adopted as opposed to the standard procedure. Also the greater proportion of error for this strategy during training is to be expected. The all-or-none strategy presents the items least likely to be in the learned state, so it is natural that more errors would be made during training.

## TEST OF A PARAMETER-DEPENDENT STRATEGY

As noted earlier, the strategy derived for the all-or-none model in the case of homogeneous items does not depend on the actual values of the model parameters. In many situations either the assumptions of the all-or-none model or the assumption of homogeneous items or both are seriously violated, so it is necessary to consider strategies based on more general models. Laubsch[4] considered the optimization

---

[4] J. H. Laubsch. An adaptive teaching system for optimal item allocation. Unpublished doctoral dissertation, Stanford University, 1969.

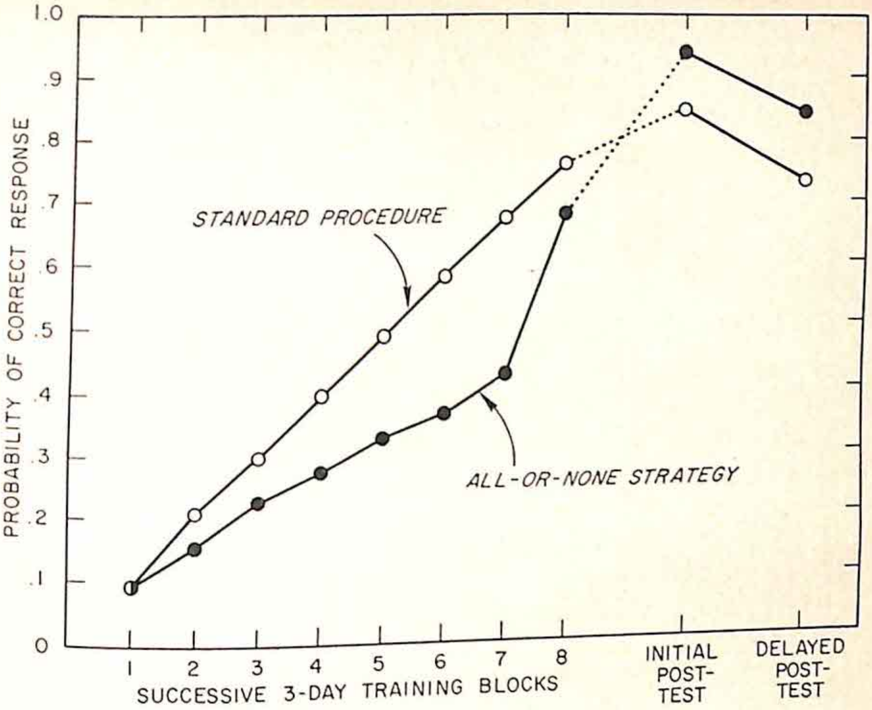


FIG. 2. Probability of correct response in Lorton's experiment.

problem for cases where the random-trial increments model is appropriate. He made what is perhaps a more significant departure from the assumptions of the all-or-none strategy by allowing the parameters of the model to vary with students and items. The following discussion is based upon Laubsch's work, but introduces a more satisfactory formulation of individual differences. This change and the estimation of initial condition parameters produce experimental measures of the effectiveness of optimization procedures that are significantly greater than those reported by Laubsch.

It is not difficult to derive an approximation to the optimal strategy for the random-trial increments model that can accommodate student and item differences in parameter values, if these parameters are known. Since parameter values must be specified in order to make the necessary calculations to determine the optimal study list, it makes little difference whether these numbers are fixed or vary with students and items. However, making estimates of these parameter values in the heterogeneous case presents some difficulties.

When the parameters of a model are homogeneous, it is possible to pool data from different subjects and items to obtain precise estimates. Estimates based on a sample of students and items can be used to predict the performance of other students or the same students on other items. When the parameters are heterogeneous, these advantages no longer exist unless variations in the parameter values take some known form. For this reason it is necessary to formulate a model stating the composition of each parameter in terms of a subject and item component.

Let $\pi_{ij}$ be a generic symbol for a parameter characterizing student $i$ and item $j$. An example of the kind of relationship desired is a fixed-effects Subjects × Items analysis of variance model:

$$E(\pi_{ij}) = m + a_i + d_j \qquad [4]$$

where $m$ is the mean, $a_i$ is the ability of student $i$, and $d_j$ is the difficulty of item $j$. Because the learning model parameters we are interested in are probabilities, the above assumption of additivity is not met; that is, there is no guarantee that Equation 4 would yield estimates bounded between 0 and 1. But there is a transformation of the parameter that circumvents this difficulty. In the present context, this transformation has an interesting intuitive justification.

Instead of thinking directly in terms of the parameter $\pi_{ij}$, it is helpful to think in terms of the odds ratio $\pi_{ij}/1-\pi_{ij}$. Allow two assumptions: (*a*) the odds ratio is proportional to student ability; (*b*) the odds ratio is inversely proportional to item difficulty. This can be expressed algebraically as

$$\frac{\pi_{ij}}{1-\pi_{ij}} = \kappa\frac{a_i}{d_j}, \qquad [5]$$

where $\kappa$ is a proportionality constant. Taking logarithms on both sides yields

$$\log\frac{\pi_{ij}}{1-\pi_{ij}} = \log\kappa + \log a_i - \log d_j. \quad [6]$$

The logarithm of the odds ratio is usually referred to as the "logit." Let $\log\kappa = \mu$, $\log a_i = A_i$, and $-\log d_j = D_j$. Then Equation 6 becomes

$$\text{logit}\ \pi_{ij} = \mu + A_i + D_j. \qquad [7]$$

Thus, the two assumptions made above lead to an additive model for the values of the parameters transformed by the logit function. Equation 7, by defining a subject-item parameter $\pi_{ij}$ in terms of a subject parameter $A_i$ applying to all items and an item parameter $D_j$ applying to all subjects, significantly reduces the number of parameters to be estimated. If there are $N$ items and $S$ subjects, then the model requires only $N+S$ parameters to specify the learning parameters for $N \times S$ subject items. More importantly, it makes it possible to predict a student's performance on items he has not been exposed to from the performance of other students on them. This formulation of learning parameters is essentially the same as the treatment of an analogous problem in item analysis given by Rasch (1966). Discussion of this and related models for problems in mental test theory is given by Birnbaum (1968).

Given data from an experiment, Equation 7 can be used to obtain reasonable parameter estimates, even though the parameters vary with students and items. The parameters $\pi_{ij}$ are first estimated for each student-item protocol, yielding a set of initial estimates. Next the logistic transformation is applied to these initial estimates, and then using these values subject and item effects ($A_i$ and $D_j$) are estimated by standard analysis of variance procedures. The estimates of student and item effects are used to adjust the estimate of each transformed student-item parameter, which in turn is transformed back to obtain the final estimate of the original student-item parameter.

The first students in an instructional program that employs a parameter-dependent optimization scheme like the one outlined above do not benefit maximally from the program's sensitivity to individual differences in students and items; the reason is that the initial parameter estimates must be based on the data from these students. As more and more students complete the program, estimates of the $D_j$s becomes more precise until finally they may be regarded as known constants of the system. When this point has been reached, the only task remaining is to estimate $A_i$ for each new student entering the program. Since the $D_j$s are known, the estimates of $\pi_{ij}$ for a new student are of the right order, although they may be systematically high or low until the student component can be accurately assessed.

Parameter-dependent optimization programs with the adaptive character just described are potentially of great importance in long-term instructional programs. Of interest here is the random-trial increments model, but the method of decomposing parameters into student and item components would apply to other models as well. We turn now to an experimental test of the adaptive optimization program based on the random-trial increments model. In this case the parameters $\alpha$, $c$, and $q'$ of the random-trial increments model were separated into item and subject components following the logic of Equation 7. That is, the parameters for subject $i$ working on item $j$ were defined as follows:

$$\begin{aligned} \text{logit}\ \alpha_{ij} &= \mu^{(\alpha)} + A_i^{(\alpha)} + D_j^{(\alpha)} \\ \text{logit}\ c_{ij} &= \mu^{(c)} + A_i^{(c)} + D_j^{(c)} \qquad [8] \\ \text{logit}\ q'_{ij} &= \mu^{(q')} + A_i^{(q')} + D_j^{(q')}. \end{aligned}$$

Note that $A_i^{(\alpha)}$, $A_i^{(c)}$, and $A_i^{(q')}$ are measures of the ability of subject $i$ and hold for all items, whereas $D_j^{(\alpha)}$, $D_j^{(c)}$, and $D_j^{(q')}$ are measures of the difficulty of item $j$ and hold for all subjects.

The instructional program was designed to teach 300 Swahili vocabulary items to college-level students. Two presentation strategies were employed: (*a*) the all-or-none procedure and (*b*) the adaptive optimization procedure based on the random-trial increments model. As in the Lorton study, a within-subjects design was employed in order to provide a sensitive comparison of the strategies. For each student two sublists of 150 items were formed at random from the master list; instruction on items from one sublist was governed by the all-or-none strategy and by the adaptive optimization strategy for the other sublist. Each day a student was tested on and studied 100 items presented in a random order; 50 items were from the all-or-none sublist chosen using the all-or-none strategy, and 50 from the adaptive optimization list selected according to that strategy. A Swahili word would be displayed, and the student was required to give its English translation. Reinforcement consisted of a printout of the correct Swahili-English pair. Twenty such training sessions were involved, each lasting for approximately 1 hour. Two or three days after the last training session an initial posttest was administered over all 300 items; a delayed posttest was given approximately 2 weeks later.

The lesson optimization program for the random-trial increments model was more complex than those described earlier. Each night the response data for that day were entered into the system and used to update parameter estimates; in this case an exact record of the complete presentation sequence and response history had to be preserved. A computer-based search algorithm was used to estimate parameters and thus the more accurate the previous day's estimates, the more rapid was the search for the updated parameter values. Once updated estimates had been obtained, they were entered into the optimization program to select individual lists for each student to be run the next day. Early in the experiment (before estimates of $D^{(\alpha)}$, $D^{(c)}$, and $D^{(q')}$ had stabilized) the computation time was fairly lengthy, but it rapidly decreased as more data accumulated, and the system homed in on precise estimates of item difficulty.

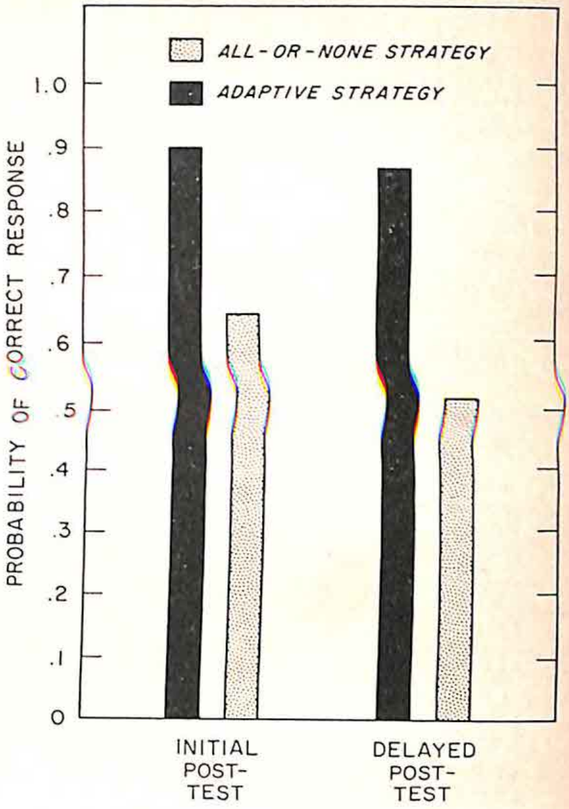


FIG. 3. Posttest performance for the all-or-none strategy and for the parameter-dependent strategy of the random-trial increments model.

Figure 3 presents the final test results and indicates that for both the initial and delayed posttests the parameter-dependent strategy of the random-trial increments model was markedly superior to the all-or-none strategy; on the initial posttest the relative improvement was 41% and 67% on the delayed posttest. It is apparent that the parameter-dependent strategy was more sensitive than the all-or-none strategy in identifying and presenting those items that would benefit most from additional study. Another feature of the experiment was that students were run in successive groups, each starting about one week after the prior group. As the theory would predict, the overall gains produced by the parameter-dependent strategy increased from one group to the next. The reason is that early in the experiment estimates of item difficulty were crude, but improved with each successive wave of students. Near the end of the experiment estimates of item difficulty were quite exact, and the only task that remained when a new student came on the system was to estimate his particular $A^{(\alpha)}$, $A^{(c)}$, and $A^{(q')}$ values.

## CONCLUDING REMARKS

The studies reported here illustrate one approach that can contribute to the development of a theory of instruction. This is not to suggest that the strategies they tested represent a complete solution to the problem of optimal item selection. The models upon which these strategies are based ignore several potentially important factors, such as short-term memory effects, interitem relationships, and motivation. Undoubtedly, strategies based on learning models that take some of these variables into account would be superior to those analyzed so far.

The task and learning models considered in this article are extremely simple and of restricted generality; nevertheless, there are at least two reasons for studying them. First, this type of task occurs in many fields of instruction and needs to be understood in its own right. No matter what the pedagogical orientation, it is hard to conceive of an initial reading program or foreign language course that does not involve some form of list-learning activity. In this respect it should be noted that the parameter-dependent strategy of the random-trial increments model has been incorporated into several parts of the Stanford computer-assisted instruction program in initial reading; although no formal evaluation has yet been made, the strategy appears to be highly effective.

There is a second reason for the type of analysis reported here. By making a careful study of a few cases that can be understood in detail, it is possible to develop prototypical procedures for analyzing more complex optimization problems. At present, analyses comparable to those reported here cannot be made for many problems of central interest to education, but by having examples of the above sort it is possible to specify with more clarity the steps involved in devising optimal procedures. Three aspects need to be emphasized: (*a*) the development of an adequate description of the learning process, (*b*) the assessment of costs and benefits associated with possible instructional actions and states of learning, and (*c*) the derivation of optimal strategies based on the goals set for the student. The examples considered here deal with each of these factors and point out the issues that arise.

It has become fashionable in recent years to criticize learning theorists for ignoring the prescriptive aspects of instruction, and some have argued that efforts devoted to the laboratory analysis of learning should be redirected to the study of learning as it occurs in real-life situations. These criticisms are not entirely unjustified for in practice psychologists have too narrowly defined the field of learning, but to focus all effort on the study of complex instructional tasks would be a mistake. Some successes might be achieved, but, in the long run, understanding complex learning situations must depend upon a detailed analysis of the elementary perceptual and cognitive processes that make up the human information handling system. The trend to press for relevance of learning theory is healthy, but if the surge in this direction goes too far, we will end up with a massive set of prescriptive rules and no theory to integrate them.

It needs to be emphasized that the interpretation of complex phenomena is problematical, even in the best of circumstances. The case of hydrodynamics is a good example for it is one of the most highly developed branches of theoretical physics. Differential equations expressing certain basic hydrodynamic relationships were formulated by Euler in the eighteenth century. Special cases of these equations sufficed to account for a wide variety of experimental data. These successes prompted Lagrange to assert that the success would be universal were it not for the difficulty in integrating Euler's equations in particular cases. Lagrange's view is still widely held, in spite of numerous experiments yielding anomalous results. Euler's equations have been integrated in many cases, and the results were found to disagree dramatically with observation, thus contradicting Lagrange's assertion. The problems involve more than mere fine points, and raise serious paradoxes when extrapolations are made from results obtained in the laboratory to actual conditions.

The following quotation from Birkhoff (1960) should strike a sympathetic chord among those trying to relate psychology and education:

> These paradoxes have been the subject of many witticisms. Thus, it has been said that in the nineteenth century, fluid dynamicists were divided into hydraulic engineers who observed what could not be explained, and mathematicians who explained things that could not be observed. It is my impression that many survivors of both species are still with us [p. 26].

Research on learning appears to be in a similar state. Educational researchers are concerned with experiments that cannot be readily interpreted in terms of learning theoretic concepts, while psychologists continue to develop theories that seem to be applicable only to the phenomena of the laboratory. Hopefully, work of the sort described here will bridge this gap and help lay the foundations for a theory of instruction.

## REFERENCES

ATKINSON, R. C. Computer-assisted learning in action. *Proceedings of the National Academy of Sciences*, 1969, **63**, 588–594.

ATKINSON, R. C., & CROTHERS, E. J. A comparison of paired-associate learning models having different acquisition and retention axioms. *Journal of Mathematical Psychology*, 1964, **2**, 285–315.

ATKINSON, R. C. & FLETCHER, J. D. Teaching children to read with a computer. *The Reading Teacher*, 1972, **25**, 319–327.

BELLMAN, R. *Adaptive control processes*. Princeton: Princeton University Press, 1961.

BIRKHOFF, G. *Hydrodynamics: A study in logic, fact, and similitude*. (2nd ed.) Princeton: Princeton University Press, 1960.

BIRNBAUM, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley, 1968.

BRUNER, J. S. Some theorems on instruction stated with reference to mathematics. In E. R. Hilgard (Ed.), *Theories of learning and instruction: The Sixty-Third Yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press, 1964.

CALFEE, R. C. The role of mathematical models in optimizing instruction. *Scientia: Revue Internationale de Synthèse Scientifique*, 1970, **105**, 1–25.

DEAR, R. E., SILBERMAN, H. F., ESTAVAN, D. P., & Atkinson, R. C. An optimal strategy for the presentation of paired-associate items. *Behavioral Science*, 1967, **12**, 1–13.

GROEN, G. J., & ATKINSON, R. C. Models for optimizing the learning process. *Psychological Bulletin*, 1966, **66**, 309–320.

HILGARD, E. R. (Ed.) *Theories of learning and instruction: The Sixty-Third Yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press, 1964.

KARUSH, W., & DEAR, R. E. Optimal stimulus presentation strategy for a stimulus sampling model of learning. *Journal of Mathematical Psychology*, 1966, **3**, 19–47.

NORMAN, M. F. Incremental learning on random trials. *Journal of Mathematical Psychology*, 1964, **2**, 336–350.

RAIFFA, H., & SCHLAIFFER, R. *Applied statistical decision theory*. Cambridge, Mass.: M.I.T. Press, 1968.

RASCH, G. An individualistic approach to item analysis. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science*. Chicago: Science Research Associates, 1966.

SMALLWOOD, R. D. The analysis of economic teaching strategies for a simple learning model. *Journal of Mathematical Psychology*, 1971, **8**, 285–301.

(Received October 15, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 62–63

# TWO INTERPRETATIONS OF BROWN'S HYPOTHESIS FOR THE RISKY SHIFT

RUSSELL D. CLARK III[1]
*Florida State University*

EDWIN P. WILLEMS
*University of Houston*

In the growing literature on the risky shift, two different interpretations of Roger Brown's risk-as-value hypothesis have been formulated. The value placed on risk is to be (*a*) more risky than other persons or (*b*) at least as risky. Most investigators have advocated the former view. However, the present empirical data provide a clear basis for choosing the at least as risky interpretation.

The most persistent and popular interpretation of the group-induced shift toward risk is Roger Brown's (1965) risk-as-value hypothesis. This view makes two assumptions: (*a*) that moderate risk is a widely held cultural value and (*b*) that this value leads individuals to want to be *at least as willing* as their peers to take risks. When a group of such persons meets and discusses matters of risk taking, some discover that their own positions, relative to those of others, are not as risky as they had thought, and they shift toward greater risk in order to achieve the relative position that they value.

Many authors interpret Brown's hypothesis to say that individuals view themselves as and desire to be *more* risky than their peers (e.g., Jellison & Riskind, 1970; Levinger & Schneider, 1969; Teger & Pruitt, 1967; Wallach & Wing, 1968). Furthermore, two major reviews of research on the risky shift (Dion, Baron, & Miller, 1970; Vinokur, 1971) both used this formulation. In contrast, Clark and Willems (Clark, 1971; Clark, Crockett, & Archer, 1971; Clark & Willems, 1969; Willems, 1969; Willems & Clark, 1969; Willems & Clark, 1971) and Vidmar (1970) are the only investigators who have stated consistently in their published papers that the cultural value of risk leads individuals to view themselves as being *at least as willing* as their peers to take risks. Wallach, one of the most frequent contributors to the area, has shifted his usage from the "more risky" interpretation (Wallach & Wing, 1968) to the "at least as risky" interpretation (Wallach & Mabli, 1970). Thus, of the three major current reviews, two (Dion et al., 1970; Vinokur, 1971) use one formulation, and one (Clark, 1971) uses the other. The purpose of this note is to highlight several arguments in favor of Clark's usage.

One step toward settling the issue is to refer to Brown (1965). In the primary statement of his risk-as-value hypothesis, Brown reviewed some unpublished data and concluded: "Not one subject guessed that these others would answer more riskily than themselves; they guessed the others to be the same as themselves or more conservative [p. 700]." Then Brown (1965) stated his risk-as-value postulate: "It looks as if each person answering as an individual conceives himself to be *at least as risky* [italics added] as the average of a reference group [p. 700]."

Brown notwithstanding, the distinction between more risky and at least as risky may appear to be a minor one. However, since it can be shown that the distinction has profound theoretical, empirical, and perhaps even practical implications, the record should be set straight.

Theoretically, the two interpretations lead to very different predictions. If the value is to be more risky, then members of groups whose risk preferences are homogeneous should shift toward greater risk because each group member would discover that he was not *riskier* than other members. If, on the other hand, the value is to be at least as risky, members of homogeneous groups should not shift toward risk because they would discover that they were already *at least as risky* as the other members. Willems and Clark (1971) found no risky shifts in homogeneous groups but did find them in heterogeneous groups. It is of interest to point out that while Willems and Clark (1971) interpreted this result as supporting Brown's (1965) hypothesis, Dion et al. (in press), because they adopted the incorrect, more risky view, interpreted these results as disconfirming the risk-as-value hypothesis.

---

[1] Requests for reprints should be sent to Russell D. Clark III, Department of Psychology, Florida State University, Tallahassee, Florida 32306.

Furthermore, if the more risky interpretation were correct, then the average risk levels in groups should escalate continuously over time, until they reached the ceiling of the risk scale, because each group member would want to remain "ahead of the pack" on risk. What we find, instead, is the kind of ceiling effect that the at least as risky interpretation would predict: The average risk level after group participation almost never exceeds the initial risk level of the riskiest member, which is usually below the highest possible level of risk (Clark, 1971).

The existing empirical data on riskiness of self in relation to others also favor the at least as risky view. We have found, across seven samples totaling approximately 700 subjects from the Midwest, Southwest, and Southeast, that a stable and predictable 70% of subjects responded to the Kogan and Wallach (1964) choice dilemmas questionnaire (the most widely used measure of risk) in terms of being at least as risky as peer groups, whereas approximately only 40% responded in terms of being more risky.

Finally, if we envision a time when the generality of the risky shift phenomenon has been established and its coordination to real-life situations has been worked out, we can foresee practical implications as well. Consider how different would be the predictions and implications for social engineering if members of juries, investment firms, gambling groups, medical diagnostic teams, fire departments, police departments, and adolescent gangs all wished to be *more* risky than their peers rather than *at least as risky*.

Until recently, the literature could accommodate both interpretations, but with the appearance of the three reviews and the study by Willems and Clark (1971), the distinction becomes more important. It should be clear that, in terms of origin and theoretical and empirical tenability, Clark is on firmer ground in his rendering of the hypothesis than are either Vinokur (1971) or Dion et al. (1970).

## REFERENCES

Brown, R. *Social psychology.* New York: The Free Press, 1965.

Clark, R. D., III. Group-induced shift toward risk: A critical appraisal. *Psychological Bulletin,* 1971, **76**, 251–270.

Clark, R. D., III, Crockett, W. H., & Archer, R. L. The relationship between perception of self, others, and the risky shift. *Journal of Personality and Social Psychology,* 1971, **20**, 425–429.

Clark, R. D., III, & Willems, E. P. Where is the risky shift? Dependence on instructions. *Journal of Personality and Social Psychology,* 1969, **13**, 215–221.

Dion, K. L., Baron, R. S., & Miller, N. Why do groups make riskier decisions than individuals? In L. Berkowitz (Ed.), *Advances in experimental social psychology.* New York: Academic Press, 1970.

Jellison, J. M., & Riskind, J. A social comparison of abilities interpretation of risk-taking behavior. *Journal of Personality and Social Psychology,* 1970, **15**, 375–390.

Kogan, N., & Wallach, M. A. *Risk taking: A study in cognition and personality.* New York: Holt, Rinehart & Winston, 1964.

Levinger, G., & Schneider, D. J. A test of the "risk is a value" hypothesis. *Journal of Personality and Social Psychology,* 1969, **11**, 165–169.

Teger, A. I., & Pruitt, D. G. Components of group risk taking. *Journal of Experimental Social Psychology,* 1967, **3**, 189–205.

Vidmar, N. Group composition and the risky shift. *Journal of Experimental Social Psychology,* 1970, **6**, 153–166.

Vinokur, A. A review and theoretical analysis of the effects of group processes upon individual and group decisions involving risk. *Psychological Bulletin,* 1971, **76**, 231–250.

Wallach, M. A., & Mabli, J. Information versus conformity in the effects of group discussion on risk taking. *Journal of Personality and Social Psychology,* 1970, **14**, 149–156.

Wallach, M. A., & Wing, C. W. Is risk a value? *Journal of Personality and Social Psychology,* 1968, **9**, 101–106.

Willems, E. P. Risk is a value. *Psychological Reports,* 1969, **24**, 81–82.

Willems, E. P., & Clark, R. D., III. Dependence of the risky shift on instructions: A replication. *Psychological Reports,* 1969, **25**, 811–814.

Willems, E. P., & Clark, R. D., III. Shift toward risk and heterogeneity of groups. *Journal of Experimental Social Psychology,* 1971, **7**, 304–312.

(Received October 30, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 64–69

# USE OF RANK ORDER DATA IN FUNCTIONAL MEASUREMENT[1]

DAVID J. WEISS[2]
*California State College, Los Angeles*

NORMAN H. ANDERSON
*University of California, San Diego*

Subjects estimated average value of angle pairs using magnitude estimations or graphic ratings. These numerical response data follow a simple averaging model. Functional scaling yields a linear relation between subjective and objective angle. The numerical data are then reduced to rank orders, and J. B. Kruskal's monotone analysis of variance (MONANOVA) procedure is applied. This allows a reconstruction of the original metric information from the strictly ordinal information, illustrating the power of MONANOVA in scaling. Limitations of MONANOVA in testing the underlying model are discussed.

Functional measurement procedures have been useful in the study of stimulus integration (e.g. Anderson, 1971; Shanteau[3]; Weiss & Anderson, 1969). The subject is required to integrate several informational stimuli into a single judgment, and a primary theoretical problem is the nature of the integration rule. The integration rule itself involves two scaling problems: measuring the subjective values of the response and also of the stimuli. Functional measurement provides a simple basis for the simultaneous solution of these three problems (Anderson, 1970, Figure 1).

All this work has used numerical response measures which, in the quantitative tests, need to be on equal interval scales. Simple rating scales have worked surprisingly well, but some response measures will not ordinarily be interval scales. Functional measurement allows for a monotone transformation to rectify the response scale (Anderson, 1962b; Bogartz & Wackwitz, 1971). This procedure utilizes whatever metric information is in the measured response.

An extreme case arises when the overt response is in the form of ranks. Of themselves, the ranks contain no metric information, and it might seem impossible to derive any. However, fundamental work by Shepard (1962, 1966) and by Kruskal (1964, 1965) has demonstrated the contrary: with sufficient constraints and assumptions, rank order techniques can, in certain respects, approach the power of interval methods. The present report supports the Shepard-Kruskal approach by demonstrating an equivalence between functional scales obtained from numerical responses and scale values derived from strictly ordinal properties of the same data.

## JUDGMENTAL TASK AND MODEL

A simple task of psychophysical integration was used: Subjects were presented with two angles and asked to estimate their average inclination. The obvious model assumes that the subjective response is just an average of the subjective values of the inclinations. With a factorial stimulus design, the overt response to the two angles in Cell $ij$ would then be

$$R_{ij} = ws_i + (1 - w)t_j \qquad [1]$$

where $s_i$ and $t_j$ are the subjective values of the row and column stimuli, and $w$ and $1 - w$ their relative weights. Unequal weighting can handle order or position effects.

Two properties of this simple model have fundamental importance (Anderson, 1970, pp. 155ff). First, if Equation 1 holds, then the data will plot as a set of parallel lines. This parallelism prediction, testable by analysis of variance, constitutes a joint test of the model and of the overt response scale. In this way, therefore, the model can provide a validational base for the response scale. Magnitude

---

[1] This work was supported by National Science Foundation Grants GB-6666 and GB-21028.

[2] We wish to thank J. B. Kruskal for making his MONANOVA program available and for his helpful comments on some of these analyses. We also wish to thank James C. Shanteau for good advice, as well as Carlyn Joergensen who ran the subjects in this experiment.

Requests for reprints should be sent to David J. Weiss, Department of Psychology, California State College, Los Angeles, California 90032.

[3] J. C. Shanteau. Component processes in risky decision judgments. Unpublished doctoral dissertation, University of California, San Diego, 1970.

estimation, as has been noted (e.g., Garner & Creelman, 1967; Shepard, 1966), lacks such a validational base. The simple psychophysical integration task, despite its apparent triviality, has a basic theoretical role to play (Weiss & Anderson, 1969). Simple integration tasks, it should be added, do not always follow simple models. Thus, Birnbaum, Parducci, and Gifford (1971) and Kreuger (1970) found discrepancies from simple models for averaging and adding two lengths, respectively.

Second, the averaging model, if successful, provides interval scales of the stimuli simply and directly. Indeed, the row means of the factorial design estimate the subjective values of the row stimuli on an equal interval scale (Anderson, 1962a).

## EXPERIMENTAL PROCEDURE

The subject was presented with cards containing two angles, constructed according to a 5 × 5 factorial design. His task was to judge their average inclination using two response modes, graphic ratings, and magnitude estimation.

*Graphic rating response.* Each response was a mark across a horizontal line, 20 millimeters below and parallel to the long edge of a 137 × 72 millimeter sheet of paper. Two anchor stimuli were present throughout this part of the experiment. One had 10-degree and 20-degree angles, the other had 160-degree and 165-degree angles. The subject was told that the anchors corresponded to positions "about here," 2 centimeters from the left and right ends of the scale, respectively.

*Magnitude estimation.* A standard stimulus with the angles 60 degrees and 120 degrees was always present in this part of the experiment. The subject was told that the average inclination of the standard was to be called 100, and that each stimulus was to be judged in terms of its ratio to the standard.

*Stimuli.* The 5 × 5 design yielded 25 angle pairs. The left angle had the values 15, 45, 75, 105, and 135 degrees; the right angle had the values 30, 60, 90, 120, and 150 degrees. Each angle was defined by two segments of 30-millimeter length, with vertex toward the left. One segment of each angle was horizontal, 42 millimeters from the lower edge of the long side of the 127 × 203 millimeter stimulus card.

*Procedure.* The data reported here are from the second and fourth sessions of a five-session experiment, the other sessions of which were concerned with similar judgments of grayness. In each session, the subject judged four replications of a group of single angles, followed by four replications of the 25 angle pairs under one response mode. The eight subjects received $1.87 per hour. Trials took approximately 15 seconds for magnitude estimation, about 5 seconds longer for the graphic rating since the experimenter read the response to the nearest millimeter on each trial.

## RESULTS

*Averaging model.* Figure 1 plots the raw data means for the two response conditions. The right-hand angle is on the horizontal axis, and the five curves correspond to values of the left-hand angle. The averaging model predicts that these curves should be parallel, and this is approximately the case. The graphic ratings are a bit more regular than the magnitude estimations, a recurrent finding in the writers' research program. The plot of geometric means of the magnitude estimations gave virtually the identical picture.

The statistical test of parallelism is given by the two-way interactions in the analysis of variance. These were nonsignificant, with $F = .91$ and $1.31$ ($df = 16/112$), for graphic and magnitude ratings, respectively. Overall, the data follow the model quite well, though deviations in particular subjects or in particular components of the interaction cannot be ruled out.

*Nonmetric rank analysis.* Since the raw data appear reasonably additive, it is appropriate to ask whether they can be reconstructed from knowledge of the rank orders alone. Accordingly, the 25 responses within each replicate for each subject were ranked and fed into Kruskal's (1965) MONANOVA computer program. This program finds that monotone transformation which, in effect, makes factorial data maximally additive. Fitted values were computed for each replicate for each subject.

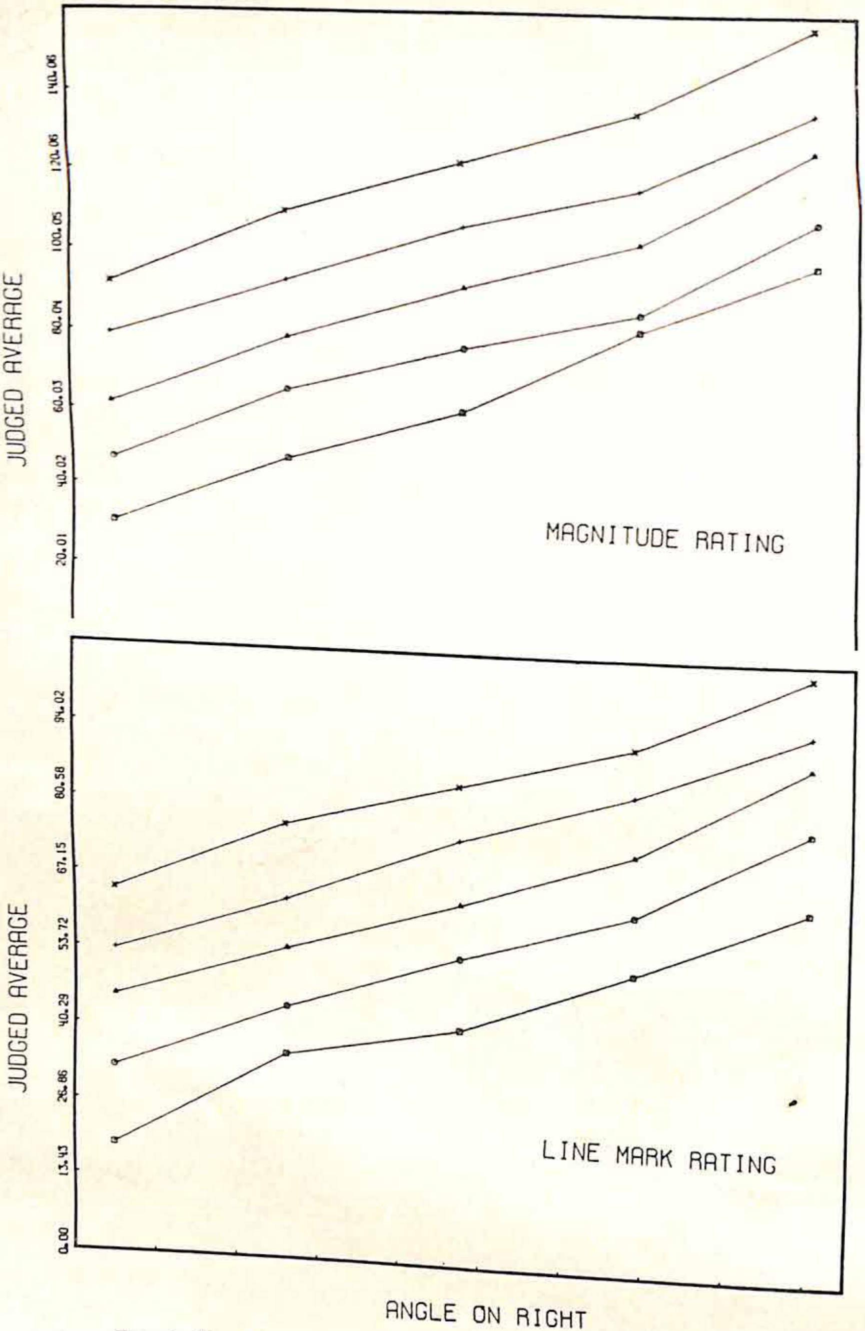


FIG. 1. Raw data means for the two response conditions.

The means of these fitted values are in Figure 2, plotted in the same manner as Figure 1. These reconstructed data appear even more parallel than the raw data, reflecting the success of the additivity-inducing transformation.

*Stimulus scaling*. According to functional measurement theory, the marginal means of the factorial design constitute interval scales of the subjective values of the stimuli. The linearity of the curves in Figure 1, and their equidistant vertical spacing, mean that the subjective and physical values of the angle stimuli are linearly related. Because of the averaging model, this linearity reflects a corresponding linear relation between the aver-

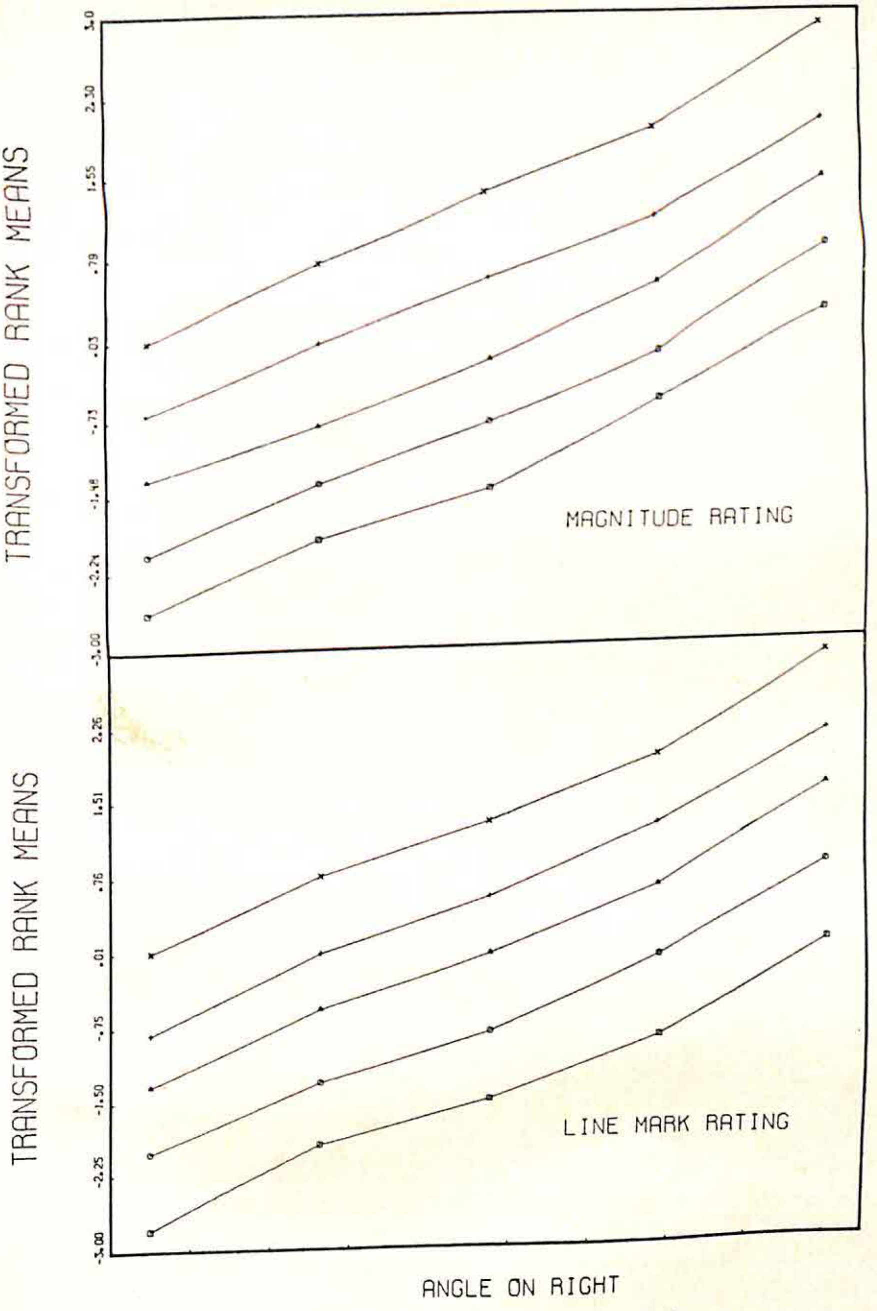


FIG. 2. Fitted data means for two response conditions.

aging response and the mean physical angle, consistent with the work of Miller and Sheldon (1969) on average inclination.

It is important to know whether the transformed ranks actually recover the original metric information. The additivity in the reconstructed data does not guarantee this with a small stimulus design. Accordingly, scale values were estimated from the reconstructed data of Figure *2* in the same way as for the original data of Figure 1. These two sets of scale values are compared in Figure 3. The left-hand angle was chosen arbitrarily, and the marginal raw means plotted on the horizontal, with the transformed rank means on the vertical. Except for a reversal in the magnitude estimation data for one subject, the curves are approximately linear. That means

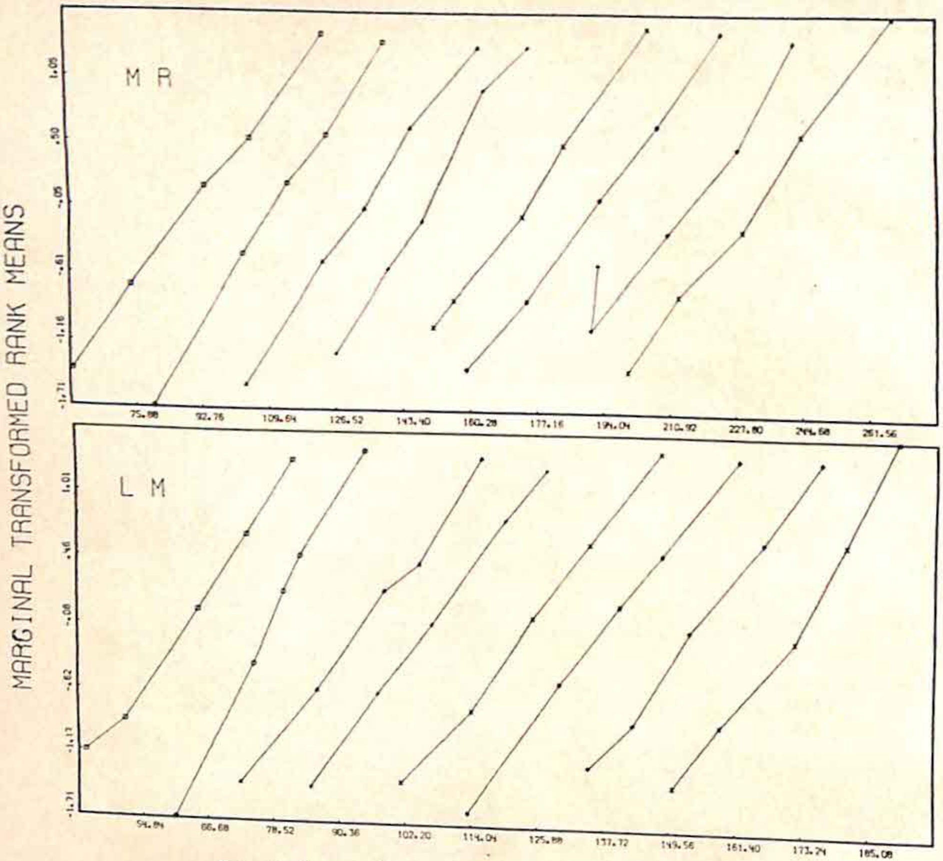


Fig. 3. Comparison of two sets of scale values.

that the rank analysis has recovered the original metric information.

## Discussion

The present results illustrate both positive and negative features of nonmetric MONANOVA analysis. On the positive side is the impressive reconstruction, from the ranks alone, of the metric information of the raw data. This is no surprise, of course, in view of the extensive work on multidimensional scaling by Shepard and by Kruskal. However, it does validate their method for a case in which the criterion was set by the initial metric analysis based on functional measurement.

This last consideration brings out certain limitations of MONANOVA, both technical and substantive. At a substantive level, the analysis is completely dependent on the assumption of additivity. If the additive model does not apply, then the MONANOVA scaling will not in general be valid; interval scaling from ordinal data requires knowledge of the appropriate model. And at the technical level, unfortunately, the nonmetric analysis makes it difficult to test whether the basic model is correct. Two main cases require brief comment.

If only rank orders are available, no reasonable assessment of response variability seems possible. With suitably spaced stimuli, for instance, different replications must be expected to yield identical rank orderings. No variability estimate is then available. More generally, any estimate of variability would depend on stimulus spacing as well as true response variability.

MONANOVA is not limited to rank order data, but may also be applied to numerical responses. A transformation based on cell means would then allow a valid estimate of within-cell response variability. However, an unknown number of degrees of freedom are used up so that a valid test would still not be available.

Valid tests could be obtained if MONANOVA were used with preliminary data to define the form of the transformation. In particular, one random half of the data could be used to define the transformation, the other half to test goodness of fit. Alternatively, a power series (Anderson, 1962b; Bogartz & Wackwitz, 1971) allows at least an approximately valid test.

In short, the MONANOVA technique is very powerful for the purpose of scaling, but only if it is known that an additive model applies. It is less effective at determining whether or not such a model does apply. This is a serious problem because there is considerable evidence for averaging processes in judgment, and averaging models are additive only under certain restrictions.

The functional measurement approach has been primarily concerned with the judgment model itself, additive or nonadditive. Scaling, both of stimulus and response, is important, but mainly because it is functional in the elucidation of the integration model.

## REFERENCES

ANDERSON, N. H. Application of an additive model to impression formation. *Science*, 1962, **138**, 817–818. (a)

ANDERSON, N. H. On the quantification of Miller's conflict theory. *Psychological Review*, 1962, **69**, 400–414. (b)

ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170.

ANDERSON, N. H. Integration theory and attitude change. *Psychological Review*, 1971, **78**, 171–206.

BIRNBAUM, M., PARDUCCI, A., & GIFFORD, R. K. Contextual effects in information integration. *Journal of Experimental Psychology*, 1971, **88**, 149–157.

BOGARTZ, R. S., & WACKWITZ, J. H. Polynomial response scaling and functional measurement. *Journal of Mathematical Psychology*, 1971, **8**, 418–443.

GARNER, W. R., & CREELMAN, C. D. Problems and methods in psychological scaling. In H. Helson & W. Bevan (Eds.), *Contemporary approaches to psychology*. Princeton, N. J.: Van Nostrand, 1967.

KREUGER, L. E. Apparent combined length of two-line and four-line sets. *Perception & Psychophysics*, 1970, **8**, 210–214.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, **29**, 1–27.

KRUSKAL, J. B. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society* (Series B), 1965, **27**, 251–263.

MILLER, A. L., & SHELDON, R. Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology*, 1969, **81**, 16–21.

SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 1962, **27**, 125–140.

SHEPARD, R. N. Metric structures in ordinal data. *Journal of Mathematical Psychology*, 1966, **3**, 287–315.

WEISS, D. J., & ANDERSON, N. H. Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, 1969, **82**, 52–63.

(Received November 9, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 1, 70–71

# HOW BLIND IS BLIND?

## A SIMPLE PROCEDURE FOR ESTIMATING OBSERVER NAIVETÉ

WILLIAM W. BEATTY

*North Dakota State University*

A simple procedure is described for estimating the naiveté ("blindness") of observers with respect to their knowledge of the experimental treatments received by subjects whose behavior they are scoring. Such information is relevant to assessing the possibility of observer bias in clinical and experimental studies where judgments of human observers represent the primary data.

Single and double blind procedures are widely employed to control for the effects of suggestibility in patients receiving pharmacological or other forms of therapy. They are also used to prevent observer biases from differentially influencing the results of studies where measurement of the dependent variable cannot be conveniently achieved by the use of automated equipment.

While the necessity for using judgments of human observers as the principal data in an investigation is probably more frequent in clinical psychology and other applied areas, many dependent variables of interest to experimental psychologists are typically measured observationally. For example, in the 1968 volumes of the *Journal of Comparative and Physiological Psychology*, 29 articles (including studies of sensitivity to electric shock, reflex fighting, imprinting, and maternal behavior) employed judgments of human observers as a primary source of data. In 22 of these studies no mention was made of whether the observers were aware of the treatments received by the subjects. Seven reports stated that the observers were blind, but in no case was any attempt made to determine whether the observers were in fact unaware of the treatments received by the subjects. Since most observational techniques require sophisticated observers, it is not unlikely that "blind" observers may discover the treatment conditions of the subjects through some leak in laboratory security. Moreover, sophisticated observers are likely to be aware of the purpose of an experiment even if they are not explicitly informed of it, and, by observation of side effects of treatments on the appearance or behavior of subjects, they may be able to identify the treatments that the subjects have received.

Since the cues that observers may make use of may be subtle and therefore difficult to identify, or unavoidable consequences of the experimental treatments, attempts to control this potential source of observer bias are unlikely to be successful. However, it is relatively simple to determine whether or not the observer is really aware of the treatments subjects have received. Merely by asking the observer to guess the treatment combinations each subject has received and by comparing the frequency of correct identifications obtained to that expected by chance, one can estimate whether or not the observers were in fact naive. If a chi-square test reveals that the observers were able to identify the treatment conditions received by subjects significantly more often than expected by chance, the possibility of observer bias should be considered in interpreting the results of the investigation.

A recent study (Beatty & Beatty, 1970) provides an illustration of this technique. In this experiment the sensitivity to electric shock of 78 female rats that received one of eight different combinations of neonatal and adult hormone treatments was measured by a modification of the Evans (1961) procedure. Although the observer was nominally blind, the hormone treatments produced changes in body weight, in phallic length, and in the texture and appearance of the animals' coats that could have provided clues for the

---

[1] Requests for reprints should be sent to William W. Beatty, Department of Psychology, North Dakota State University, Fargo, North Dakota 58102.

identification of the treatments that the subjects received. To assess this possibility, the observer was asked to guess the exact combination of treatments each subject received. By chance alone the observer would be expected to correctly identify the subjects' experimental condition 1 out of 8 times in general, or 9.75 times in this case. In fact, the observer guessed correctly 18 times (23%). Thus in this case the observer correctly identified the treatments received by subjects significantly more often than would have been expected by chance ($\chi^2 = 7.97$, $df = 1$, $p < .01$), and the possibility that observer bias influenced the results must be considered.

While measurement of the [illegible]o which observer naiveté has been ac[illegible] not solve the problem of observer b[illegible]d provide information that is useful for the evaluation of studies using observational techniques. The desirability of achieving more objective measurements of dependent variables is indisputable, but many variables of considerable interest to psychologists cannot be measured except by human observers. The simplicity of the present technique should make it attractive to workers who depend on observational measurements.

A similar approach could be used to determine the degree of naiveté of patients in pharmacological studies. This procedure would seem useful in such studies, since many drugs produce changes in autonomic functions (e.g., heart rate, blood pressure) that do not result from placebo treatment. Such changes could alert the patient that he is receiving the experimental treatment, thus reducing the effectiveness of placebo controls for suggestibility.

## REFERENCES

Beatty, W. W., & Beatty, P. A. Hormonal determinants of sex differences in avoidance behavior and reactivity to electric shock in the rat. *Journal of Comparative and Physiological Psychology*, 1970, 73, 446–455.

Evans, W. O. A new technique for the investigation of some analgesic drugs on a reflexive behavior in the rat. *Psychopharmacologia*, 1961, 2, 318–325.

(Received November 12, 1970)

---

## Manuscripts Accepted for Publication in the *Psychological Bulletin*

Comment on "Linguistic Deficiency and Thinking: Research with Deaf Subjects 1964–1969." Harry Bornstein (Department of Psychology, Gallaudet College, Kendall Green, Washington, D. C. 20002) and Howard L. Roy.

The Devil Rides Again: Correlation as an Index of Fit. Michael H. Birnbaum (Department of Psychology, University of California at Los Angeles, Los Angeles, California 90024).

## NOTICE TO AUTHORS

In response to a request from the APA Publications Board, the selected reviewers of *Psychological Bulletin* have agreed to participate in a blind reviewing system. An effort will be made to insure that the individuals who review manuscripts will have no information regarding the identity of the author or of the institution with which the author is affiliated.

Authors submitting manuscripts to *Psychological Bulletin* are requested to include with each copy of the manuscript a cover sheet, which contains the title of the manuscript, the name of the author or authors, and the author's institutional affiliation. The date the manuscript is submitted also should appear on the cover sheet. The first page of the manuscript should not contain the author's name or institutional affiliation and should contain the title of the manuscript and the date it is submitted. Footnotes containing information pertaining to the identity of the author or his institutional affiliation should be on separate pages. Every effort should be made by the author to see that the manuscript itself contains no clues as to his identity.

The selected reviewers are not aware that knowledge of the author's identity previously has influenced their judgments, but they recognize this as a possibility, and the blind review system therefore is being initiated.

Authors are requested to use the proposed format immediately, and manuscripts received after July 1, 1972, that do not conform to these suggestions will be returned to the author.

Date 42
Lib

VOL. 78, NO. 2 AUGUST 1972

Psychological Bulletin

# EGOCENTRISM AND SOCIAL INTERACTION ACROSS THE LIFE SPAN[1]

WILLIAM R. LOOFT[2]

*University of Wisconsin*

Existing theory and data pertaining to the concept of egocentrism are organized and interpreted within a life-span framework. Most research and theorization on egocentrism to date are found to be concerned with only the infancy and childhood periods of life, although some theorization has concerned the adolescent period. Essentially none of the existing literature pertaining to the major portion of the life span—adulthood and senescence—is found to be directly related to the concept of egocentrism. However, the notions of rigidity and regression and the process of psychosocial disengagement, all of which have been traditionally identified with the aging process, are reinterpreted as either manifestations of or contributors to increasing egocentricity in later life. Several areas of pertinent research are identified.

> *Foregoing self, the universe grows I.*
> Edwin Arnold

It is the position of this article that our understanding of human behavior and development will be considerably advanced when research is conceptualized, conducted, and interpreted within the frame of reference of the life span, thereby taking into account the continuous processes of change that characterize all behavior. By no means is the present author the first to call attention to the merits of the life-span approach (cf. Bayley, 1963; Pressey & Kuhlen, 1957), but few systematic attempts at formulating theory and research within this framework have been produced to date. The recent West Virginia Conference on Life-Span Developmental Psychology was one notable endeavor intended to stimulate work in this area (see Goulet & Baltes, 1970; Hooper, 1970). The present article is intended to provide a conceptual framework for further systematic efforts in life-span psychology.

Specifically, this article uses the concept of *egocentrism* as the focal element for organizing and interpreting existing theory and research in a life-span framework. A strong case can be made regarding the influence of egocentrism in the lives of members of the human species, both individually and collectively. Piaget (1962) has emphasized that a major portion of the developmental energies within the individual is invested into the process of overcoming his egocentric reference system. Piaget has even further suggested that egocentricity of thought—the illusions caused by the immediate point of view—has perhaps been the central problem in the history of human affairs. An attempt is made here to extend the egocentrism concept, which heretofore has been associated primarily with the early years of life, to adulthood and senescence, and to present a framework for interpreting the manner by which an individual's lifelong interaction in the social sphere modifies both his cognitive and affective modes of functioning.

## THE CONCEPT OF EGOCENTRISM

Egocentrism has been a central concept in cognitive developmental psychology. The term

---

[1] The preparation of this paper was facilitated by a grant from the Graduate School of the University of Wisconsin.

[2] Requests for reprints should be sent to William R. Looft, who is now at the College of Human Development, Pennsylvania State University, University Park, Pennsylvania 16802.

© 1972 by the American Psychological Association, Inc.

Piaget (1928) put it this way:

> What then gives rise to the need for verification? Surely it must be the shock of our thought coming into contact with that of others, which produces doubt and the desire to prove. . . . The social need to share the thought of others and to communicate our own with success is at the root of our need for verification [p. 204].

It is clear, therefore, that within the Piagetian system the occurrence of communication conflicts is a necessary condition for intellectual decentration. However, the statements above suggest that the remission of egocentrism is more than just a cognitive phenomenon. The disequilibrium initiated by conflicts in social exchange has an affective component as well as the generally recognized cognitive component. As Langer (1969a) has pointed out, change seems to occur only when the child *feels*, consciously or unconsciously, that something is wrong. Therefore, both affective and organizational (intellectual) disequilibrium are necessary conditions for the child's engagement in the adaptive, constructive mental activities that comprise the sources for cognitive change and development.[5]

Feffer (1959, 1970) has provided a detailed theoretical statement that has important implications for the egocentrism/social interaction issue. His position proposes that the decentering concept provides continuity between internal cognitive organization and the cognitive organization and the cognitive structuring of social content. This extension of the decentering concept to interpersonal behavior rests on the assumption that the same processes, or at least parallel processes, are involved in both impersonal cognition (person to object) and interpersonal cognition (person to person) (Feffer, 1970).

Furthermore, according to Feffer, this decentering interpretation also can be applied to the structuring of events at different levels of cognitive maturity. The argument is as follows: The dovetailing of responses involved in effective social interaction demands that each participant modify his intended behavior in anticipation of the other's reaction to this behavior. In order to anticipate accurately this reaction, one must be able to see his intended behavior from the perspective of the other person, but at the same time he must also view his intended action from his own perspective. The cognitive organization of an individual capable of effective social interaction can be interpreted as one in which different perspectives are considered simultaneously in relation to each other such that the distortion engendered by a given viewpoint (or centering) is equilibrated or corrected by another perspective. In contrast, a child who is able to focus only sequentially upon his behavior from a single perspective at a time will have difficulty in appropriately modifying his responses.

To summarize, the significance of social life for cognitive development rests on the conceptualization of social interaction as a necessary condition for transition from one developmental level to another. The essential difference between the Piagetian view of the value of social interaction and the social learning view is that for the former, social contacts create stress and conflict that induce cognitive transformation; for the latter, social interactions provide a system of reinforcement for certain behaviors leading toward growth.

The remainder of this article reviews pertinent literature within the egocentrism/social interaction framework as it relates to traditionally defined age segments of the life span. Only a brief overview of the research on children and adolescents is presented, since this material has been extensively reviewed by previous writers. Emphasis is placed on the adult and aged portions of the life span.

## EGOCENTRISM AND SOCIAL INTERACTION IN INFANCY AND CHILDHOOD

Childhood egocentrism assumes different forms as the child progresses to higher levels of cognitive maturity. Inhelder and Piaget (1958) have delineated three types of egocentrism within this period.

1. At the sensorimotor level (0–2 years) the child's interactions are characterized by a radical egocentrism—a complete lack of differentiation between himself and his actions and the characteristics of the given situation. The developmental landmark of this stage, which occurs with the first decentering process,

---

[5] Langer (1969a) has provided an excellent discussion of the notion of disequilibrium as a source for developmental change.

is the practical dissociation of these poles and the consequent formation of the object concept and the differentiation of his own ego.

2. The preoperational child (2–7 years) displays a new capacity for representation (which was the element that released the infant from object egocentrism), but it is this element that now ensnares him in egocentricity with respect to symbols. This type is especially evident in the child's linguistic behavior, for he seems to assume that words carry much more information than they actually do. Thus a lack of differentiation exists between the child's own viewpoint and those of other persons. The breakdown of representational egocentrism occurs with the emergence of concrete operational thought. At this time the child gains facility in dealing with two elements or relations simultaneously, and therefore he can retain both the symbol and its referent in his mind at the same time and thus distinguish between them.

3. The third form of egocentrism is evident in school-age children as they acquire concrete operational thought. At this time, when the child is mastering classes, relations, and quantities, egocentrism is manifested in his inability to differentiate accurately between mental products (resulting from his newly acquired operations) and perceptual givens. When the school child constructs a hypothesis, he naturally assumes this is a product imposed by the external data rather than a result of his own thought. If he is challenged, typically the child will reinterpret that data to be congruent with this hypothesis, rather than change his initial stance. With the emergence of formal operational thought in early adolescence, this form of egocentricity eventually breaks down. The person discovers the arbitrariness of his mental constructions and learns to separate them from perceptual phenomena. (Another form of egocentrism, which appears with the onset of adolescence, is discussed in a subsequent section.)

As implied in the discussion above, the transition from one form of egocentrism to another takes place in a dialectic fashion in such a manner that the cognitive structures that free the child from a lower form of egocentrism are the same structures that entangle him in a higher egocentric form. From a developmental perspective it seems legitimate to regard egocentrism as a negative by-product of any emergent mental system, for it produces a new set of unrealistic, nonobjective representations of the world at each newly achieved level of thought.

### *Egocentric Speech*

A primary manifestation of child's egocentrism is found in his language behavior. The concept of egocentric, or "private," speech has been the subject of a considerable amount of theorization and research; the discussion here only adumbrates this phenomenon.

Piaget (1926) categorized children's spontaneous speech productions into two major classes (and, as is his wont, into several subclasses within each major class). "Egocentric speech" is speech that whether uttered in solitude or in the presence of others, is judged to lack a primary communicative aim. There seems to be no attempt on the part of the child to take the role of the listener and hence to adapt the message to the listener's informational needs or input capacity. Indeed, there is no attempt even to make sure the listener is listening. Piaget stressed that while the young child's egocentric speech reflects a lack of cognitive skill, it more fundamentally reflects a lack of "social will" to communicate. "Socialized speech," on the other hand, applies to utterances that do seem to reflect a genuine communicative orientation. This communication is characterized by the child's very real attempts to inform the listener, to persuade or to coerce him to some course of action. Piaget further postulated that extensive social interaction, especially with peers, is the necessary element to bring about the displacement of egocentric speech with socialized speech.

Not unlike many other childhood phenomena that Piaget described, the notion of egocentric speech has become a focal point for empirical verification or refutation by other researchers. Considerable controversy continues to characterize this issue. Much of the controversy stems from definitional concerns. Alternative views or expanded delineations of private speech have been offered by Vygotsky (1962), Werner and Kaplan (1963), Flavell (1966a), and Kohlberg, Yaeger, and Hjertholm (1968). Furth (1969) has contributed to the controversy by essentially dismissing the concept altogether. He asserted that terms such as egocentric, private, and inner speech have been used with vague and unanalyzed meaning. While acknowledging that many children talk to themselves in noncommunicative situations, he nevertheless insisted that the social use of language is always foremost. "No normal child at any period of development uses language primarily for purposes of self-communication [Furth, 1969, p. 118]."

Other major controversies have concerned the frequency and the function of these speech forms. The high frequencies of egocentric productions reported by Piaget (40%–70% of the utterances of children aged 5–6 years) have been refuted by some (e.g., McCarthy, 1930) and supported by others (e.g., Smith, 1935; Weir, 1962). Disagreements over the function or form of private speech have been raised by Vygotsky (1962), Flavell and associates (Flavell, 1966a; Flavell, Beach, & Chinsky, 1966; Flavell, Botkin, Fry, Wright, & Jarvis, 1968), and Kohlberg et al. (1968). Kohlberg et al. (1968) have presented a detailed discussion of theoretical issues related to this issue and also have hypothesized a developmental hierarchy containing five levels of progression; this hierarchy represents an integration of the theoretical formulations of Piaget (1926), Mead (1934), and Vygotsky (1962).

It would appear that future research in this area will pay close attention to situational determinants of language behavior (Cazden, 1970; Moscovici, 1967). Both the frequency and the form of a child's egocentric speech productions appear to be especially susceptible to the degree of familiarity with other persons present, the social context, and the nature by which he is observed.

### *Investigations of Communicative and Interpersonal Skills*

The theoretical notions concerning egocentrism, social interaction, and related concepts in children have generated a spate of empirical investigations. The categorical names under which these research efforts fall include studies of empathic ability, role-taking skills, and verbal communication effectiveness, to name a few. In an early review of the role-taking-skills literature, Sarbin (1954) attempted to identify the essential components of interpersonal behavioral phenomena. In accordance with Piaget and others, Sarbin concluded that the essential element in any sort of skill sequence in role taking seems to be a process by which the individual somehow apprehends certain attributes of the other individual. Of necessity these must be inferential attributes (rather than directly perceptible attributes): needs, intentions, opinions, capacities, limitations, and so on. Sarbin proposed that the estimate of these attributes formed by the role taker or the effective communicator must be a synthesis of two sources of information: (*a*) knowledge of people and their behavior in various situations and (*b*) perceptual input from the overt behavior of the other person or from other cue sources in the immediate situation. However, Brown (1965), in his encoding–decoding analysis of interpersonal communication, insisted that the ability to control the lexicon and the grammar of a language is a third necessary component.

Only a few of the major areas of research endeavor related to this topic are touched upon here. For a more extensive discussion of this research, the reader is referred to Flavell et al. (1968).

An extensive program of research on role-taking and communication skills in children has been conducted by Flavell and associates (Flavell, 1966; Flavell et al., 1968; Fry, 1966, 1969). This research has systematically confirmed and extended Piaget's studies of social communication in children. Most of these studies involve some variation of a basic procedure in which different-aged children are presented with tasks requiring an ability to shift perspectives. Perhaps the principal finding, at least in relation to the present discussion, was an indication that young children, when speaking, confuse their own perspective with that of the auditor in communication situations. This confusion was found to decline regularly with age. Fry (1969), however, was generally unsuccessful in attempting to train children to provide appropriate communications to listeners with varying informational needs.

In addition to the tasks used by Flavell et al., ingenious experimental procedures have been devised by other investigators to study the egocentrism/role-taking issue. Cowan's (1966) task, for example, obtained measures of cognitive egocentrism for each child on a task adapted from Piaget's "three mountain problem." Subjects were then paired for participation in a social interaction task in which only oral cues (the subjects could not see each other) were allowed in order to achieve a common goal. From this task were obtained measures of communication efficiency, egocentric language, and the strategies of the communication system established by the participants. Performance in this situation by nonegocentric pairs has been found to be far superior to that of egocentric or mixed pairs. Cowan concluded that these data suggest that the notion of egocentric languages provides a useful conceptual tool for analyzing the link between the structural properties of the individual child and the formal properties of his social interaction.

A widely used procedure for studying referential communication is the "stack-the-blocks" game (Glucksberg, Krauss, & Weisberg, 1966). One child, designated as "speaker" and seated opposite a "listener" he cannot see, must communicate to his partner how to stack a series of wooden blocks, each of which depicts a different visual design. Performance on this task has been found to improve with increasing age and with increasing degree of familiarity of the figures being described. Studies employing this procedure with children include Glucksberg et al. (1966), Glucksberg and Krauss (1967), Krauss and Rotter (1968), and Krauss and Glucksberg (1969).

A final procedure to be described here is Feffer's (1959) Role Taking Task. This task, which was designed to relate Piaget's decentering concept to the interpersonal domain, requires the child to make up stories for a number of ambiguous scenes depicting social situations. Each scene is presented again, and the child is asked to retell the original story from the viewpoint of each of his characters. Evaluation is based on the degree to which he is able to refocus upon his initial story from the perspectives of each of his characters while maintaining continuity among the various versions. Feffer and Gourevitch (1960) found the expected improvement in Role Taking Task performance with increasing age.

Other studies investigating various aspects of role-taking disposition and referential communication abilities in children include Alvy (1968), Burns and Cavey (1957), Cohen and Klein (1968), Elkind (1961), Flapan (1968), and Sullivan and Hunt (1967).

## EGOCENTRISM AND SOCIAL INTERACTION IN ADOLESCENCE

Adolescence in cognitive developmental theory is marked by the acquisition of formal operational thought. This mode of thinking is characterized by two major abilities, which are (*a*) the ability to think about one's own thinking and (*b*) the ability to recognize possibilities as well as actualities. Yet these same higher forms of thought, which were instrumental in eliminating the egocentricity of concrete operational thought, now immerse the individual in new forms of egocentricity (Elkind, 1967; Piaget, 1967). Parallel to his ability to conceptualize his own thought, the person now also has the ability to conceptualize the thought of others. According to Elkind (1967), "It is this capacity to take account of other people's thought, however, which is the crux of adolescent egocentrism [p. 1029]." Thus, the adolescent fails to differentiate between the events to which his own thought is directed and those about which others are concerned. Undoubtedly the rapid physical and physiological changes occurring within him contribute to his self-concern; psychoanalytically oriented writers have long noted the resurgence of narcissism and self-absorption in the adolescent period (cf. Blos, 1962; Sullivan, 1953). Furthermore, the adolescent assumes that others are as obsessed with his outward appearance and behavior as he is himself. As opposed to the child's inability to take the other person's point of view, the adolescent "takes the other person's point of view to an extreme degree [Elkind, 1968, p. 153]." Elkind (1967) has proposed two concepts to describe the unique forms that adolescent egocentrism may take. The adolescent's *imaginary audience* accounts for his self-consciousness, and his *personal fable* stems from his belief in the uniqueness of his own experiences.

These forms of adolescent egocentrism do not, in the main, characterize adult thought, which suggests that some mechanism operates to dispel them in later adolescence. In this theoretical framework, it would appear that formal operational thought itself, in conjunction with the adolescent's interaction in a social world, leads to the dissolution of his egocentrism. When the adolescent thinks about his own thoughts, almost as if these thoughts were objects, he becomes able to recognize the difference between his own preoccupations and concerns and the concerns of other people. Social interaction plays a major role in this process. According to Piaget (1967):

> Adolescent social interaction . . . is aimed primarily at discussion. Whether in twosomes or in small coteries, the world is reconstructed in common, and the adolescent loses himself in endless discussion as a means of combating the real world [p. 68].

Thus, adolescent egocentrism can be seen to be overcome on two dimensions. Cognitively, this process occurs as a result of the eventual differentiation between one's own preoccupations and those of others. Affectively, this occurs as the individual gradually recognizes and integrates the feelings of others with his own feelings.

Little actual research has been conducted with adolescents that relates to this egocentrism/social interaction framework. An exception is a study by Rockway (1969), which investigated the development of the logical nature of interpersonal judgments in adolescent boys. In a task that required subjects to predict the behavior of a person in a programmed case history, Rockway noted an increase in the hypothetico-deductive quality of the boys' social judgments and a decrease in egocentric self-focusing sets. Also discerned was a trend in twelfth graders toward another form of self-focusing judgments, which Rockway labeled "enlightened egocentrism." As opposed to the blanket assumption by younger adolescents that one's own reactions are necessarily relevant data, these older adolescents sometimes focused upon their own past reactions to predict another person's behavior where it appeared logical to assume that oneself and others were in some way alike.

## EGOCENTRISM AND SOCIAL INTERACTION IN ADULTHOOD

As is the case for many forms of psychological functioning, egocentrism has not been studied in adults. Indeed, within the existing notions of cognitive development it is not even considered to be a concept relevant to adulthood. In his discussion of cognitive developmental theorists (primarily Piaget and Werner), Langer (1969b) lamented that his presentation necessarily had to end with adolescence because these theorists "have no clear conception of adult cognition that is qualitatively different from adolescent cognition, although they would hesitate to equate the two [p. 112]."

Flavell (1970), arguing from a position that developmental phenomena are inevitable, are irreversible, and have directional and uniform intraspecific characteristics (and therefore imply an underlying biological growth process), suggested the most changes found in adulthood are not truly developmental:

> There seems to be no compelling and consistent evidence that at least the sorts of intellectual changes commonly studied are inevitable, of striking magnitude, or uniform across adulthoods; it could be conceded perhaps that they are irreversible and directional, although the directionality does not seem to manifest itself in a fixed sequence of well-demarcated stages or developmental landmarks [p. 249].

Flavell further suggested that most adult cognitive changes are social interpersonal ones, due primarily to experiential factors. Thus, the sum of adult experiences results in changes in the individual's implicit theories concerning oneself, others, and the human condition in general.

Neugarten has studied extensively the kinds of social and personal changes in adulthood to which Flavell referred. In her interview studies of middle-aged men (see Neugarten, 1966), the most salient characteristic found was increased attention and concern with health problems and energy conservation. Another finding was a changing time perspective as individuals moved from young adulthood to middle adulthood. A person's orientation to life was found to shift from the perspective of time since birth to that of time left to live. Birren (1969), in his interviews with men and women aged 30–60 years, obtained many reports of reduced energy for carrying out the tasks associated with this time of life. These persons seemed to be aware that when they were younger, they reacted differently to situations and people in general.

The picture that emerges from the interview studies by Neugarten and Birren is one of a change from active to passive modes of mastering one's environment. The change is from an "active-assimilatory" style to a "passive-accommodatory" style, or a movement of energy from an outer world to an inner world orientation. Neugarten's data suggest that the sixth decade of life is typically the period when this passive and introspective characteristic markedly increases in predominance. It is important to recognize that not all middle-aged persons manifest this trend, and therefore this phenomenon would not qualify as a truly developmental one, according to the strict criteria delineated by Flavell and others (see also Kohlberg, 1969). Nevertheless, this tendency in many middle-aged people toward increased contemplation of one's inner thoughts and diminished reactivity to other persons seems related to egocentrism as conceptualized here: indeed, these persons may manifest many of the functional properties of egocentric behavior and speech in studies designed to test this proposition.

A review of existing studies of egocentric manifestations in the adulthood years is of necessity brief, for there are few studies that even indirectly relate to this question. (Of course, it is well to recognize that to date there exist very few psychological investigations of any kind of the adult years, with the important exception of the college adult population.) Seemingly egocentric behaviors in adults would most likely be found in their social interactions with others, particularly verbal communication. A review of a few studies that relate to this question are now presented.

Soskin and John (1963) analyzed a series of ordinary adult conversations. Their data suggested that private speech, as described earlier in this article, is a rare or useless category of adult speech. However, apparently no systematic studies of egocentric or private speech in adulthood have been carried out.

Shneidman and Farberow (1957), in their investigation of both successful and unsuccessful suicide attempts, asked psychologists expert in language phenomena to distinguish genuine suicide notes from simulated ones (written by nonsuicidal types asked specifically to write suicide notes). This separation was accomplished very easily. It seemed that the pretender did not always write to the point of view of his supposed addressee, and this inappropriate decentering easily revealed his identity. Typically he seemed to be writing to the informational requirements of the psychologists who were to be reading his note, not to his soon-to-be-bereaved relatives.

Werner and Kaplan (1963) instructed adults to write two sets of descriptions of novel drawings to be used at a future time. One set was written for the subject's own use, and the other was written by the subject to be used by someone else. The former descriptions were about half as long as the latter. It was apparent that the subject was addressing himself to two different sets of informational requirements. For himself he could be more brief and could perhaps make reference to idiosyncratic experience. For the unknown reader of his note he had to describe the object in greater detail, using commonly recognized referents. This factor is revealed also in the frequent occasion of an undergraduate student asking to borrow the lecture notes of another. The often-heard reply of the lender in this situation is something like, "You can borrow my notes, but I don't think you'll be able to get much out of them."

Krauss, Vivekananthan, and Weinheimer (1968) described a similar experiment. College students were engaged in a color-coding task; coding was done both for personal use and for the use of others. An analysis of the products revealed that the nonsocial coding condition produced more low frequency (i.e., unusual) words than the social condition.

Krauss and Weinheimer (1964), using the stack-the-blocks game described earlier, found that college students were able to develop standardized nomenclatures for novel patterns quite readily; children were found to have great difficulty with the same task The stack-the-blocks game appears to represent an excellent example of what is crucial to life-span research: a task that is applicable to subjects across a major portion of the age continuum. A recent study by Krauss and Glucksberg (1969) demonstrated the high utility of this procedure for comparing groups of subjects from early childhood to young adulthood. Feasibly the same procedure could be used also to investigate the decentering abilities of middle- and old-aged subjects.

It should be noted in passing that Maclay and Newman (1960) devised a similar task for adult subjects that antedates Krauss and Glucksberg's procedure. Their method required one person to describe a picture to a listener seated behind a screen so that the latter could identify it from an array of pictures before him. The results suggested that normal adult communicators are likely to show a fair amount of sensitivity to the listener's apparent input needs, augmenting their messages whenever they receive cues (such as negative feedback) that suggest that these needs are high.

The Role Taking Task of Feffer also appears to be another procedure that has wide applicability in studying decentering abilities in variously aged subjects. In addition to its use with children, the Role Taking Task has also been used as a basis for pairing college-aged subjects for participation in a social interaction task (Feffer & Suchotliff, 1966). This latter task was a Password-type game, in which each member of the pair was required to communicate test words via one-word association clues to his partner, who did not know the words. The pairs who scored high on the Role Taking Task communicated words more quickly and with fewer clues than did the lower scoring pairs.

To conclude, the picture that is suggested by a review of existing theory and research is that egocentrism, as a characteristic of mental and behavioral life, is not a salient aspect of the early and middle years of adulthood. Given the structural postulates of currently popular cognitive developmental theories, this conclusion appropriately follows. Nevertheless, convincing, if not plentiful, evidence exists that some behaviors befitting the egocentric label are manifested by some persons on some occasions or in certain contexts, suggesting that the decentering process sometimes fails, or perhaps never even takes place, in persons possessing the theoretically most advanced forms of thought. Indeed, this possibility has been recognized by no less an authority than Piaget himself:

> To take an example from adult life, every beginning instructor discovers sooner or later that his first lectures were incomprehensible because he was talking to himself, so to say, mindful of only his own point of view. He realizes only gradually and with difficulty that it is not easy to place oneself in the shoes of students who do not yet know what he knows about the subject matter of the course [Piaget, 1962, p. 5].

## EGOCENTRISM AND SOCIAL INTERACTION IN SENESCENCE

Although the underlying mechanisms are not at all well understood, extensive documentation exists that with advancing age many kinds of performance and interest changes occur.[6] As yet there is no theory extant that is able to provide a convincing account for these changes.

The points of departure for investigating the psychological aspects of the aging process have been extremely diverse, and a partial list of typical hypotheses tested in these researches illustrates the diversity of approach: brain damage, interference of previous learning, rigidity, lowered arousal potential, psychosocial disengagement, increased neural noise, and many more. As Birren (1970) pointed out, to some extent all these hypotheses may be partially correct in accounting for behavioral changes associated with aging, since they tend to apply only to the functioning of a subsystem of the organism. Nevertheless, this highly diffused approach toward the study of aging has resulted in a mélange of findings that are confusing and often contradictory. Carp (1969) has attributed much of the blame for this problem to the tendency in gerontological research to use imprecisely defined concepts (e.g., "disengagement"). This tendency allows for the possibility that different studies deal with different traits under the same name. It seems equally likely that many of the different trait names refer to the same underlying process. The present proposal to introduce the egocentrism concept as a framework for aging research is hopefully not the addition of yet one more label to the clutter of labels already found in the gerontological literature. To the contrary, the proposal asserts that many of these existing labels might be better interpreted as forms of egocentrism, thus achieving some sense of coordination and unity.

---

[6] See Birren (1964), Botwinick (1967), and Bromley (1966) for representative surveys of this literature.

This concept, however, at least in the Genevan sense of the word, is simply not to be found in the gerontological literature. The present author's perusal of several texts and other sorts of compendiums on old age revealed only a single reference to "egocentrism." Cameron (1967) investigated the relationship between "introversion" and "egocentrism" in an aged sample, but his definition of the latter term was "having to do with one's wants and needs [p. 465]," which is a rather different meaning than that employed here.

Not surprisingly, there has been little research with adults and older persons using Piagetian tasks. This undoubtedly is due, at least in part, to the prevailing notion that the kinds of mental processes with which Piaget is most concerned are invariant and irreversible. However, Piagetian-adult research efforts are growing in number, and the findings to date have been intriguing. Piaget and Inhelder (1962) made brief mention of a study in Geneva by de Ajuriaguerra, Müller, and Mandl in which a series of the traditional conservation tasks were given to a group of aged adults. They found evidence that these abilities disintegrate in reverse order of their formation in childhood (substance, weight, volume). Another study found striking decrement in performance on a conservation-of-surface task in an aged group, as compared to two groups of younger adults (Sanders, Laurendeau, & Bergeron, 1966). These two findings, particularly the former, are highly suggestive of structural changes in cognitive abilities in old age.[7]

The following discussion attempts to reinterpret certain other areas of gerontological theory and research according to the egocentrism/social interaction schema used in this paper.

## *Rigidity in the Aged*

One concept that has long been associated with old age, and one that might beneficially be interpreted as a form of egocentrism, is that of *rigidity*. Unfortunately, little commonality exists among the definitions proposed by those who have explored the dimensions of this trait. The result has been a confusing array of research methodologies and data interpretations. Of the population of definitions of rigidity, the following is a representative sample: the inability to change one's set when the objective conditions demand it (Rokeach, 1948); adherence to present performance in an inadequate way (Goldstein, 1943); the presence of strong boundaries between mental functions (Kounin, 1941); a lack of change of behavior, where a change is necessary for success at the task, and where the subject knows that a change is likely to be demanded (Chown, 1961); a tendency to perseverate and resist conceptual change, to resist acquisition of new patterns of behavior, to refuse to relinquish old and established behavior patterns (Schaie, 1958).

The typical research procedure for studying rigidity in old adults has been to administer an instrument purporting to measure this trait (such as the test of behavioral rigidity; Schaie, 1955) and obtain correlations between these scores and performance on various kinds of problem-solving tasks. As Chown (1961) pointed out, however, there are over 50 tests that have been used by different investigators as measures of rigidity. The relationship among these tests is largely unknown. Perhaps this confusing situation could be clarified somewhat by recognizing the distinction between what McAndrew (1948) has called *structural* and *behavioral* forms of rigidity. The former refers to the individual's internal

---

Also notable is a recent theoretical article by Hooper, Fitzgerald, and Papalia (1971), in which intellectual abilities of aged persons are considered within a framework integrating Piagetian theory and the fluid versus crystallized intelligence distinction.

[7] Other investigations (all Genevan) that have administered Piagetian-type concept tasks to aged subjects include de Ajuriaguerra, Boehme, Richard, Sinclair, and Tissot (1967), de Ajuriaguerra, Kluser, Velghe, and Tissot (1965), de Ajuriaguerra, Richard, Rodriquez, and Tissot (1966), and Sinclair-de-Zwart, de Ajuriaguerra, Boehme, and Tissot (1966).

rearrangement of experience and is closely related to the cognitive tasks investigated with children; the latter concerns the subject's interaction with his environment and is more related to tasks used with mature and aged persons.

A meaningful area of research may be to investigate the possibility that many forms of behavior heretofore said to reflect "rigid" response patterns in the subject might be more usefully interpreted as the subject's inability to decenter his perceptions. For example, in one report of a study of the relationship between rigidity and problem solving in an aged sample the following observation was made: "Overcoming the influence of a problem-solving set involves the ability to change one's approach or point of view [Heglin, 1956, p. 311]." This proposed parallelism, or perhaps even identity, of rigidity and the inability to center is as yet, of course, only a verbal linkage; the empirical verification remains to be established. A start in this direction would be to compare elderly subjects' performance on one or more tests of rigidity, such as Schaie's (1955), and their performance on various tests of decentering skills, such as the Role Taking Task (Feffer, 1959) and the stack-the-blocks game (Glucksberg et al., 1966).

### *Disengagement Process*

A proposition regarding the aging process that has attained considerable popularity during the past decade is the *disengagement theory*. As proposed by Cumming and Henry (1961), the theory holds that older individuals tend to dissociate themselves from some aspects of their environments, particularly those aspects necessitating social interaction. The individual is conceived as being the center of a network of social interactions; as he ages, his social life constricts, and he experiences a curtailment of involvement in social systems. This process is seen to be inevitable, universal, willingly accepted by the individual, and beneficial to both the individual and the society. Social gerontologists have not wholeheartedly endorsed the disengagement theory, especially following the spate of contradictory evidence that has been gathered since the theory's initial presentation, but its general notions seem to have served the important function of generating more discussion and research than any other theory of aging in recent years.

It is not in keeping with the purpose of the present article to present a detailed review and critique of disengagement theory and the research it has stimulated. Nevertheless, a general overview of current thinking regarding the theory's status is warranted, given the premise of this particular section of the article, that is, that disengagement is a contributing factor to egocentric behavior manifested in old age.

Maddox (1964) outlined three key ideas represented by the theory of disengagement: (*a*) the process of social and psychological withdrawal is modal for the aging population; (*b*) this process is both intrinsic (i.e., developmental) and inevitable; and (*c*) the withdrawal process is not only a correlate of successful aging but also probably a condition of it. The first point seems to have been largely verified (Maddox, 1964), but the nature of the withdrawal is by no means uniform for all elderly persons (cf. Carp, 1968; Palmore, 1968; Tallmer & Kutner, 1969). The second point is still a controversial one; apparently the condition that the process is a truly intrinsic one is open to question (Havighurst, Neugarten, & Tobin, 1968). Point *c* also remains controversial, for evidence exists that disengagement is associated with unsuccessful aging as well as the reverse (Lowenthal & Boler, 1965; Palmore, 1968). The disengagement concept appears to epitomize the problem in much of the gerontological research in that it was originally imprecisely defined, and therefore it has fostered many different studies dealing with many forms of behavior, all of which have been labeled "disengagement." As Carp (1969) has pointed out, this tendency has resulted in a mass of findings that are repetitive or contradictory. Despite these problems, the disengagement notion perists as a valuable heuristic framework for formulating research programs of psychological and sociological aspects of aging.

Much of the research concerned with the disengagement question has produced evidence that is highly revealing with regard to the changing nature of the elderly person's interactions with other persons. A troublesome question, as yet not satisfactorily answered, is whether the decrease in social interaction among the aged is a cause or a result of other phenomena associated with aging. Havighurst et al. (1968) concluded from their studies that the individual's withdrawal has intrinsic, or developmental, qualities as well as responsive ones. Furthermore, their data indicated that as social withdrawal proceeds, an accompanying development is an increasing preoccupation with the self and a decreasing investment of emotional energy in other persons in the environment.

The studies by Birren and his colleagues (Birren, 1970) suggest that the aging individual may begin to employ a different kind of cognitive strategy or tactic in his dealings with other persons. Interviews with physicians and lawyers in later middle-age revealed that they strategically protect themselves from both cognitive and affective overload by not becoming as involved with their clients as they were accustomed to in their earlier years of practice. It is significant, perhaps, that these people were cognizant of the changing nature of their interactions with others.

Further evidence pertinent to the thesis of this article comes from a series of studies by Riegel (1972). Subjects were asked to list the names of persons they had met during their lifetime. Riegel found that young and middle-aged adults demonstrated a recency effect, that is, the greatest number of persons recalled were these encountered in the recent past. In contrast, the old-adult sample (mean age 73) did not evidence this recency effect; rather, a primacy effect was most apparent. Riegel concluded that this latter finding suggests one or both of two possibilities: (*a*) the older person encounters fewer people in his senescence; or (*b*) the people he meets in old age are of little importance to him at this time, at least in comparison to those met earlier in life. These data suggest that the older individual lives most intensely with his remote past; persons most recently encountered are of lesser significance. A recent report by Pressey and Pressey (1969) gives credence also to the first of Riegel's suggested possibilities. After retirement from active professional life, the Presseys moved into a home for the elderly. Careful frequency counts of persons encountered and interacted with were kept both before and after their retirement. As soon as 3 months after formal retirement they found that their number of social contacts was only 25% of that experienced immediately before retirement.

It must be recognized, of course, that the environments in which aged persons live are not identical, and different environments tend to possess varying degrees of social contacts. Rosow (1967), for example, found that in those age-segregated communities or cities where the density of elderly people is very high, the social interaction of the aged is fairly high. On the other hand, extensive evidence exists indicating that in age-segregated institutions and homes, there is likely to be relatively little social interaction among the aged (cf. Lieberman, 1969).

In addition to sociological observations and explanations of disengagement, it is feasible that future research will discover that the apparent lack of involvement of older persons in the social world may have to do with conservation of energy. Psychophysiological evidence indicates that certain forms of psychological stress produce longer periods of autonomic reactivity in older persons than they do in younger persons (Powell, Eisdorfer, & Bogdanoff, 1964). The speculation follows that "since the homeostatic mechanism in the aged may be more vulnerable, . . . it may simply be physiologically more sound for him to remain uninvolved [Eisdorfer, 1970, p. 66]." Sherwood (1970) has further proposed the intriguing hypothesis that social interaction in the aged is functionally related to nutritional status. This proposal follows from the consideration that a lack of companionship and sharply curtailed social interaction may be considered by many elderly people to be a state of "deprivation," thereby causing emotional stress. Emotional stress, in turn, is believed to have negative nutritional effects, especially on nitrogen balance.

From several sources, therefore, there is evidence that the world of the aging individual is one of constriction of social experience and reduced contact and perhaps even interest in other persons. It is not clear at this time just what is cause and what is effect, or whether the presumed disengagement is the result of the general social structure of society, willful intent of the individual, gradual behavioral and interest changes due to fundamental physiological responsiveness, or the product of all these factors in addition to other factors as yet unidentified. In any case, what is important to the present thesis is that the individual tends to disengage from the larger social world with advancing age. This circumstance of diminishing social interaction may contribute to a deterioration of the ability to decenter the perspective of oneself from those of others.

### *Regression as a Developmental Process of Aging*

An analysis of the older person's social world suggests the likelihood that egocentric behaviors will increase during this time. Furthermore, the process would appear to be the reverse of the progression displayed by children as they overcome their egocentric representations of reality, due to increased peer–peer interactions, which provide opportunities for practice in communication and for verification of one's perspective. Conversely, as in the case of the aged, diminished social interaction of necessity implies diminished social feedback.

The theoretical framework within which to account for changes in the individual's cognitive functioning as he gradually socially disengages during later life is not totally clear at this time. Perhaps a starting point is the account provided by Hunt (1961), which outlined the process involving the interaction of environmental circumstances with developmental functions in children, following Piaget's general formulations.[8] According to Hunt's argument, the greater the variety of situations to which the child must accommodate his "behavioral structures," the more differentiated and mobile these structures become. Extrapolating these notions to the circumstances of the aged, it would follow that the well-documented process of restriction of one's environment (due to infirmity, expectations, or whatever) eventuates in less differentiated and less mobile cognitive structures.

The idea of developmental *regression* is not a new one. Several years ago Dennis (1948) hypothesized that with the onset of old age there occurs a regression to childlike conceptual thinking. While the regression hypothesis has not been a popular one, Dennis' data, particularly those indicating the reappearance of animistic thought in elderly people, were suggestive of such a process. The notion of "regression in the service of the ego" has long been accepted in psychoanalysis, and it has been characterized as a very common process in the aged (Zinberg, 1963). However, this form of regression—primarily a greater acceptance of dependency—does not appear to imply permanent cognitive change, for it is manifested most often only by ill or exceedingly aged persons. Thus, the psychoanalytic phenomenon of regression seems to serve a functional yet temporary purpose, very much like the "regression" in moral development noted in certain college students (Kohlberg & Kramer, 1969). In a field-theory approach to describing behavioral changes with age, Schaie (1962) has also hypothesized the regression phenomenon to account for increasingly inappropriate responses of the elderly person as a result of his progressively constricting life space. The life space of the aged adult, Schaie postulated, assumes an increasing similarity with the constricted environment of the child.

Before the proposal of the postulation of the regression process to account for increased egocentricity in advanced age proceeds further, several caveats are in order. Thus far the regression hypothesis has been used as a heuristic device with respect to certain formal, organizational features of behavior in old age. It is important that such an account should not stop with the establishment of similarities among certain developmental series. As Wapner (1969) has forcefully pointed out, "*Similarity is not identity.* Ultimately, the problem is one of *establishing differences as well as similarities* [p. 41]." To understand cognitive development, it is important that the analysis not stop with the examination of the *content* of responses; more importantly, the *operations* generating the responses are primary in order of importance. It would seem that discussions of regression in the aged, such as those mentioned above, have ignored this distinction between content and operation; the prevailing assumption seems to have been that responses resembling earlier modes of thought indeed reflect the use of that earlier mode of thought. In the present discussion of the recession of egocentricity in early life and its likely reappearance in later life, it is well to heed the advice of Turiel (1969) by recognizing that "the conditions causing regressive change may not be the same as those causing progressive change [p. 113]."

---

[8] It appears that Ausubel's (1958) formulations anticipated much of Hunt's (1961) thesis.

Acknowledging these several qualifications, it would seem that regression as a developmental process in old age remains as a viable theoretical possibility. Langer (1969b) has recently made a strong case for the regression concept. The important distinction that Langer has made is that regression is always a change to a lower order stage as it exists presently in the individual's functioning, not to a stage as it existed previously. When higher structures deteriorate, the tendency is to use the next lowest functional structures available. Thus, according to Langer, the developmental direction in aging is backward, but it is backward to the forms of functional structures and actions that are present; the regression is never back to earlier, childhood forms of operation. An older person may give responses in a testing situation (such as a test of conservation of surface) that are very much like those of a child in the same situation, but these responses, though similar in content, may be the product of quite different cognitive operations.

Regression to lower order, currently existing cognitive structures, a process encouraged perhaps by a constricting social environment, may therefore account for the appearance of egocentric behaviors in persons of advanced age. Kohlberg and Kramer (1969) have suggested that such backward developments are a rare category of behaviors. However, a reinterpretation of many forms of data on aged adults, such as increasing rigidity, within the egocentrism/social interaction schema developed here, suggests the potentiality of the regression phenomenon.

## *Role of Egocentrism in Understanding the Aging Process*

Is egocentrism, therefore, a viable construct in understanding the behavior of aged adults? A cautionary note about the "aged" as a population is in order, since enormous diversity exists within this group. It is unlikely that a single process of aging will ever be found to account for the wide variety of aging patterns that exist. Nevertheless, the well-documented deterioration of ability and performance in a wide variety of psychological tasks seems to be found nearly universally in older people. Schaie and Strother (1968) found that even in a group of highly educated and financially secure older persons, considerable decrement existed in psychomotor speed, memory, and motor-cognitive flexibility. Of course, there are great individual differences in the onset of these forms of decrement.

Presently there is little direct empirical support for the notion of egocentrism as manifested in the behavior of the elderly. However, one directly relevant study (Looft & Charles, 1971) used Cowan's (1966) social interaction paradigm with adults. An impersonal, cognitive measure of decentering ability was first obtained from each subject in young-adult and old-adult samples. Subjects within each age group were then paired together to participate in the social interaction task, used previously only with children, that was described earlier in this paper (Cowan, 1966). Young adults were found to be markedly less egocentric than the old adults in the impersonal task, but in the social interaction task there was no significant difference between age groups in mean performance accuracy (though the young adults were somewhat superior). However, the mean time required to complete the social task by the older pairs was more than twice that required by the younger pairs. This latter finding could be accounted for by the response style among the aged known as "cautiousness" (Botwinick, 1969), which has been described as either a purposeful slowdown in responding for the purpose of increasing accuracy in performance or as a defensive reluctance to venture a response for fear of recognizing one's inadequacy. An additional finding of interest was that subjects' performance in the impersonal task was unrelated to performance in the social task, in both age groups; this finding is in striking contrast to Cowan's earlier studies with children. This study suggests that an adult relies extensively on the feedback he receives from other persons. In social interaction situations, whether in real life or in tests devised by psychologists, verbal interaction quickly brings into sharp focus any conflict in viewpoints, and the resulting interchange resolves the disparity. This reasoning would further suggest that when old adults are removed from social contexts, thereby removing the support of verbal interchange and correction, more primitive modes of functioning may be allowed to operate. Young adults, with apparently greater ability to decenter in impersonal contexts, are not thus hindered.

An incidental finding in the Looft and Charles study gave further indication of deficient decentering skill in aged adults. Flavell (1966b) and Glucksberg and Krauss (1967) have commented on the "verbal-pointing" phenomenon in young children when addressing a listener not visible. This involves making gestures in the air with one's hand in order to describe an object to the unseen listener and giving vague, nondirectional instructions. Several such examples of verbal pointing were found in the old-adult pairs in the present study; none was observed among the young adults.

The usefulness of the egocentrism/social interaction notion in accounting for many of the behaviors manifested by aging persons cannot be further evaluated until considerably more directly pertinent data are gathered. At this point, nevertheless, it is well to acknowledge that individual who appears to have recognized long ago this process of aging at work. In the autumn of his life, G. Stanley Hall made the following observation:

> Now that the pressure of outer reality and its duties remit, attention tends more to focus on self, and introspontaneity and mentation may take on a slightly dreamy character in that it is less under the domination of the objective environment, from which there is a new sense of freedom. The demand for rigorous proof of one's theorizations is somewhat less insistent and critics of them are felt to be lacking in insight. . . . and as the senses begin to grow dim their verification of our speculations seems a trifle less imperative [Hall, 1922, p. 401].

## CONCLUSIONS

Hopefully the present use of egocentrism as a unifying concept to account for many of the individual's behaviors across his life span amounts to substantially more than strands of gossamer strung wistfully about largely unrelated fragments of theory and data. Further work is undeniably needed to gain certainty as to the validity of this approach. Presently considerable data exist that attest to the presence and influence of egocentrism in infancy and childhood. (It is well to point out, however, that notions of egocentrism in infants are founded, to a great degree, upon extrapolations backward from observations made on older children's behavior, and thus are highly inferential.) As for the adolescent period of life, appealing theoretical accounts for adolescent egocentric behaviors have been developed, but to date little empirical testing of these notions has occurred. Clearly the greatest need for further work is left for the vast portion of the life span—the adult years. Admittedly little psychological investigation of any kind has been concerned with this age group; the need for developmental research here is clear. Finally, there is a growing body of theory and data on the later years of life (certainly much more than that which concerns the middle years of life), but little of it is directly applicable to the question of egocentrism and social interaction. This article has attempted to point out that much of the existing literature on aging can usefully be reinterpreted within this framework.

To speak of egocentrism in a life-span orientation necessitates the consideration of the distinction between structure and content properties of behavior. Egocentric behaviors in the earliest years of life are assumed to be directly attributable to the structural properties of the person's cognition, according to the theory of Piaget and its derivatives. Egocentrism in adolescence also reflects structural characteristics, given the Piagetian assertion that egocentric behaviors are negative byproducts of newly emerging forms of thought. However, if egocentriclike behaviors were noted in early and middle adulthood, nonstructural explanations for their existence would have to be nominated; existing cognitive developmental theory does not allow for the possibility of egocentric behaviors as a direct product of the structural properties of adult cognition. Apparently these behaviors would be accounted for by assuming that they are serving some temporary, regressive function for the individual, perhaps to satisfy certain affective needs. On the other hand, the absence of egocentric manifestations in adulthood would be completely consistent with this theoretical paradigm. Egocentric behaviors in senescence seemingly present a more complex problem, for on the practical side, it is likely that far more egocentriclike behaviors are to be found at this time than during middle adulthood, yet current cognitive developmental theories do not comfortably allow for their occurrence. Further research should begin to provide answers regarding the structural properties of cognitive operations at this time of life. Langer's (1969b) recent formulations appear to represent a major theoretical development to guide this needed research.

Several lines of research are suggested to begin answering questions raised in the present article:

1. A valuable first step would be to gather data describing the egocentric content of speech from the time the infant begins his first speech productions until advanced old age. Context variables will have to be manipulated carefully in such research (e.g., in solitude, in social dyads and in larger groups, with strangers, with familiar persons, etc.), and reliable rating scales for determining the egocentricity of each speech production will be needed. Flavell et al. (1968) and Kohlberg et al. (1968) have made notable progress in developing such scales.

2. Vital research will concern the relationship among measures of egocentrism or decentering ability and many other presumably related measures, including perceptual skills, various subtests in adult scales of intelligence, rigidity, and indexes of psychosocial disengagement.[9] The application of research designs incorporating multivariate longitudinal and cross-sectional sequences, as developed by Baltes and Nesselroade (1970), appears to represent an ideal approach to this area of study. These strategies would result in comparative factor analyses for evaluating the effects of age and generational changes on these behavioral variables in terms of both factor loading patterns and factor scores.

3. Recognition of the possibility of sex differences in decentering skills seems important. These differences may be especially notable in later adulthood and old age. Lowenthal (1965) has found that an important factor in satisfactory adjustment in later life is the presence of a "confidant," or someone with whom one can talk out problems and other concerns. The stereotype of the "real man," particularly among lower-class males, which holds that "a real man can handle his problems by himself" and thus belittles the need for a confidant, suggests the possibility that certain males in our population may experience even greater decline in decentering skills as an eventual result of their self-imposed avoidance of important kinds of social interaction.

4. A greater exploration of the relation between cognitive skills and affectivity at all points of the life span is warranted. Elkind (1967) and Feffer (1967) have suggested that egocentrism and decentering skills may represent valuable starting points in reconciling cognitive structure and personality dynamics. Piaget (1962) has also suggested that many forms of primitive interpersonal feelings may be due to systematic errors in the individual's emotional perspective. The Role Taking Task of Feffer (1959) represents a useful vehicle for research in this area of study.

5. The assumption that childhood egocentrism possesses the developmental criterion of universality needs to be better established. Greenfield and Bruner (1969) have proposed that egocentrism may be only a relative phenomenon due to cultural conditions and values. Their studies of children in societies encouraging individual value orientations (such as the United States) and in societies foster-

---

[9] In an exploratory study with children, Shantz examined the convergent validity of three measures related to the egocentrism trait: spatial egocentrism, communicative egocentrism, and role-taking ability. With no controls made for age or IQ, the intercorrelations among the three measures ranged only from −.03 to .18. (C. U. Shantz. Egocentrism in children: Its generality and correlates. Unpublished manuscript, Merrill-Palmer Institute, Detroit, Michigan, 1968.)

ing collective value orientations (such as the Wolof people of Senegal) produced markedly different findings. Further cross-cultural work with basic tests of decentering skill are therefore highly encouraged.

6. Further development of measures of egocentrism and decentering ability is of great priority. Measures that are applicable to persons across wide portions of the life span are especially desirable.

## REFERENCES

Ajuriaguerra, J. de, Boehme, M., Richard, J., Sinclair, H., & Tissot, R. Désintégration des notions de temps dans les démences dégénératives due grand âge. *Encephale,* 1967, **5**, 385–438.

Ajuriaguerra, J. de, Kluser, J., Velghe, J., & Tissot, R. Praxies ideatoires et permanence de d'objet. Quelques aspects de leur désintégration conjointe dans les syndromes démentiels du grand âge. *Psychiatrie Neurologique,* 1965, **150**, 306–319.

Ajuriaguerra, J. de, Richard, J., Rodriquez, R., & Tissot, R. Quelques aspects de la désintégration des praxies ideomatrices dans les démences du grand âge. *Cortex Cerebral,* 1966, **2**, 438–462.

Alvy, K. T. Relation of age to children's egocentric and cooperative communication. *Journal of Genetic Psychology,* 1968, **112**, 275–286.

Ausubel, D. P. *Theory and problems of child development.* New York: Grune & Stratton, 1958.

Baltes, P. B., & Nesselroade, J. R. Multivariate longitudinal and cross-sectional sequences for analyzing ontogenetic and generational change: A methodological note. *Developmental Psychology,* 1970, **2**, 163–168.

Bayley, N. The life span as a frame of reference in psychological research. *Vita Humana,* 1963, **6**, 125–139.

Birren, J. E. *The psychology of aging.* Englewood Cliffs, N. J.: Prentice-Hall, 1964.

Birren, J. E. Age and decision strategies. In A. T. Welford & J. E. Birren (Eds.), *Decision making and age.* Basel: S. Karger, 1969.

Birren, J. E. Toward an experimental psychology of aging. *American Psychologist* 1970, **25**, 124–135.

Blos, P. *On adolescence: A psychoanalytic interpretation.* New York: Free Press, 1962.

Botwinick, J. *Cognitive processes in maturity and old age.* New York: Springer, 1967.

Botwinick, J. Disinclination to venture response versus cautiousness in responding: Age differences. *Journal of Genetic Psychology,* 1969, **115**, 55–62.

Bromley, D. B. *The psychology of human ageing.* Baltimore: Penguin, 1966.

Brown, R. *Social psychology.* New York: Free Press, 1965.

Burns, N., & Cavey, L. Age differences in empathic ability among children. *Canadian Journal of Psychology,* 1957, **11**, 227–230.

Cameron, P. Introversion and egocentricity among the aged. *Journal of Gerontology,* 1967, **22**, 465–468.

Carp, F. M. Some components of disengagement. *Journal of Gerontology,* 1968, **23**, 382–386.

Carp, F. M. Compound criteria in gerontological research. *Journal of Gerontology,* 1969, **24**, 341–347.

Cazden, C. B. The situation: A neglected source of social class differences in language use. *Journal of Social Issues,* 1970, **26**(2), 35–60.

Chown, S. M. Age and the rigidities. *Journal of Gerontology,* 1961, **16**, 353–362.

Church, J. *Language and the discovery of reality.* New York: Vintage, 1961.

Cohen, B. D., & Klein, J. F. Referent communication in school age children. *Child Development,* 1968, **39**, 597–609.

Cowan, P. A. Cognitive egocentrism and social interaction in children. *American Psychologist,* 1966, **21**, 623 (Abstract)

Cumming, E., & Henry, W. E. *Growing old.* New York: Basic Books, 1961.

Dennis, W. Regression to childhood concepts in old age. *American Psychologist,* 1948, **3**, 308. (Abstract)

Eisdorfer, C. The implications of research for medical practice. *Gerontologist,* 1970, **10** (1, Pt. II), 62–67.

Elkind, D. The child's conception of right and left: Piaget replication study IV. *Journal of Genetic Psychology,* 1961, **99**, 269–276.

Elkind, D. Egocentrism in adolescence. *Child Development,* 1967, **38**, 1025–1034.

Elkind, D. Adolescent cognitive development. In J. F. Adams (Ed.), *Understanding adolescence.* Boston: Allyn & Bacon, 1968.

Elkind, D., & Scott, L. Studies in perceptual development: I. The decentering of perception. *Child Development,* 1962, **33**, 619–630.

Feffer, M. The cognitive implications of role-taking behavior. *Journal of Personality,* 1959, **27**, 152–168.

Feffer, M. Symptom expression as a form of primitive decentering. *Psychological Review,* 1967, **74**, 16–28.

Feffer, M. A developmental analysis of interpersonal behavior. *Psychological Review,* 1970, **77**, 197–214.

Feffer, M. H., & Gourevitch, V. Cognitive aspects of role-taking in children. *Journal of Personality,* 1960, **28**, 383–396.

Feffer, M. H., & Suchotliff, L. Decentering implications of social interaction. *Journal of Personality and Social Psychology,* 1966, **4**, 415–422.

Flapan, D. *Children's understanding of social interaction.* New York: Teachers College Press, 1968.

Flavell, J. H. *The developmental psychology of Jean Piaget.* Princeton, N. J.: Van Nostrand, 1963.

Flavell, J. H. Le langage prive. *Bulletin de Psychologie,* 1966, **19**, 698–701. (a)

Flavell, J. H. Role-taking and communication skills in children. *Young Children,* 1966, **21**, 164–177. (b)

Flavell, J. H. Cognitive changes in adulthood. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Theory and research.* New York: Academic Press, 1970.

Flavell, J. H., Beach, D. R., & Chinsky, J. M. Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development,* 1966, **37,** 283–299.

Flavell, J. H., Botkin, P. T., Fry, C. L., Wright, J. W., & Jarvis, P. E. *The development of role-taking and communication skills in children.* New York: Wiley, 1968.

Freud, A. *Normality and pathology in childhood: Assessments of development.* New York: International Universities Press, 1965.

Fry, C. L. Training children to communicate to listeners. *Child Development,* 1966, **37,** 675–685.

Fry, C. L. Training children to communicate with listeners who have varying listener requirements. *Journal of Genetic Psychology,* 1969, **114,** 153–166.

Furth, H. G. *Piaget and knowledge: Theoretical foundations.* Englewood Cliffs, N. J.: Prentice-Hall, 1969.

Glucksberg, S., & Krauss, R. M. What do people say after they have learned how to talk? Studies of the development of referential communication. *Merrill-Palmer Quarterly,* 1967, **13,** 309–316.

Glucksberg, S., Krauss, R. M., & Weisberg, R. Referential communication in nursery school children: Method and some preliminary findings. *Journal of Experimental Child Psychology,* 1966, **3,** 333–342.

Goldstein, K. Concerning rigidity. *Character and Personality,* 1943, **11,** 209–226.

Goulet, L. R., & Baltes, P. B. (Eds.) *Life-span developmental psychology: Theory and research.* New York: Academic Press, 1970.

Greenfield, P. M., & Bruner, J. S. Culture and cognitive growth. In D. A. Goslin (Ed.), *Handbook of socialization theory and research.* Chicago: Rand McNally, 1969.

Hall, G. S. *Senescence.* New York: Appleton, 1922.

Havighurst, R. J., Neugarten, B. L., & Tobin, S. S. Disengagement and patterns of aging. In B. L. Neugarten (Ed.), *Middle age and aging.* Chicago: University of Chicago Press, 1968.

Heglin, H. J. Problem solving set in different age groups. *Journal of Gerontology,* 1956, **11,** 310–317.

Hooper, F. H. The West Virginia conference on life-span developmental psychology. *Human Development,* 1970, **13,** 53–60.

Hooper, F. H., Fitzgerald, J., & Papalia, D. Piagetian theory and the aging process: Extensions and speculations. *Aging and Human Development,* 1971, 2, 3–20.

Houssiadas, L., & Brown, L. B. The coordination of perspectives by mentally defective children. *Journal of Genetic Psychology,* 1967, **110,** 211–215.

Hunt, J. McV. *Intelligence and experience.* New York: Ronald Press, 1961.

Inhelder, B., & Piaget, J. *The growth of logical thinking from childhood to adolescence.* New York: Basic Books, 1958.

Kohlberg, L. Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research.* Chicago: Rand McNally, 1969.

Kohlberg, L., & Kramer, R. Continuities and discontinuities in childhood and adult moral development. *Human Development,* 1969, **12,** 93–120.

Kohlberg, L., Yaeger, J., & Hjertholm, E. The development of private speech: Four studies and a review of theories. *Child Development,* 1968, **39,** 691–736.

Kounin, J. S. Experimental studies of rigidity. *Character and Personality,* 1941, **9,** 251–282.

Krauss, R. M., & Glucksberg, S. The development of communication: Competence as a function of age. *Child Development,* 1969, **40,** 255–266.

Krauss, R. M., & Rotter, G. S. Communication abilities of children as a function of status and age. *Merrill-Palmer Quarterly,* 1968, **14,** 161–173.

Krauss, R. M., Vivekananthan, P. S., & Weinheimer, S. "Inner speech" and "external speech": Characteristics and communication effectiveness of social and nonsocially encoded messages. *Journal of Personality and Social Psychology,* 1968, **9,** 295–300.

Krauss, R. M., & Weinheimer, S. Changes in the length of reference phrases as a function of social interaction: A preliminary study. *Psychonomic Science,* 1964, **1,** 113–114.

Langer, J. Disequilibrium as a source of development. In P. Mussen, J. Langer, & M. Covington (Eds.), *Trends and issues in developmental psychology.* New York: Holt, Rinehart & Winston, 1969. (a)

Langer, J. *Theories of development.* New York: Holt, Rinehart & Winston, 1969. (b)

Lieberman, M. A. Institutionalization of the aged: Effects on behavior. *Journal of Gerontology,* 1969, **24,** 330–340.

Looft, W. R., & Charles, D. C. Egocentrism and social interaction in young and old adults. *Aging and Human Development,* 1971, 2, 21–28.

Lowenthal, M. F. Antecedents of isolation and mental illness in old age. *Archives of General Psychiatry,* 1965, **12,** 245–254.

Lowenthal, M. F., & Boler, D. Voluntary versus involuntary social withdrawal. *Journal of Gerontology,* 1965, **20,** 363–371.

Maclay, H., & Newman, S. Two variables affecting the message in communication. In D. K. Wilner (Ed.), *Decisions, values and groups.* New York: Pergamon, 1960.

Maddox, G. L. Disengagement theory: A critical evaluation. *Gerontologist,* 1964, **4,** 80–82.

McAndrew, H. Rigidity and isolation: A study of the deaf and the blind. *Journal of Abnormal and Social Psychology,* 1948, **43,** 476–494.

McCarthy, D. *The language development of the preschool child.* (Institute of Child Welfare Monograph, Serial No. 4) Minneapolis: University of Minnesota Press, 1930.

Mead, G. H. *Mind, self, and society.* Chicago: University of Chicago Press, 1934.

Moscovici, S. Communication processes and the properties of language. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 3. New York: Academic Press, 1967.

Neugarten, B. L. Adult personality: A developmental view. *Human Development,* 1966, **9**, 61–73.

Palmore, E. B. The effect of aging on activities and attitudes. *Gerontologist,* 1968, **8**, 259–263.

Piaget, J. *Language and thought of the child.* London: Routledge, Kegan Paul, 1926.

Piaget, J. *Judgment and reasoning in the child.* New York: Harcourt, Brace, 1928.

Piaget, J. *The psychology of intelligence.* New York: Harcourt, Brace, 1950.

Piaget, J. Comments on Vygotsky's critical remarks concerning *The language and thought of the child,* and *Judgment and reasoning in the child.* Attachment to L. S. Vygotsky, *Thought and language.* Cambridge, Mass.: M. I. T. Press, 1962.

Piaget, J. *Six psychological studies.* New York: Random House, 1967.

Piaget, J., & Inhelder, B. *Le développement des quantités physiques chez l'enfant.* Paris: Delachaux & Niestlé, 1962.

Piaget, J., & Inhelder, B. *The psychology of the child.* New York: Basic Books, 1969.

Powell, A. H., Eisdorfer, C., & Bogdonoff, M. D. Physiological response patterns observed in a learning task. *Archives of General Psychiatry,* 1964, **10**, 192–195.

Pressey, S. L., & Kuhlen, R. G. *Psychological development through the life span.* New York: Harper, 1957.

Pressey, S. L., & Pressey, A. D. "Insider" longitudinal evaluations of institutional living. *Proceedings of the 77th Annual Convention of the American Psychological Association,* 1969, **4**, 457–458. (Summary)

Riegel, K. F. Time and change in the development of the individual and society. In H. W. Reese (Ed.), *Advances in child development and behavior.* Vol. 7. New York: Academic Press, 1972, in press.

Rockway, A. M. Cognitive factors in adolescent person perception development. *Developmental Psychology,* 1969, **1**, 630.

Rokeach, M. Generalized rigidity as a factor of ethnocentrism. *Journal of Abnormal and Social Psychology,* 1948, **43**, 259–277.

Rosow, I. *Social integration of the aged.* New York: Free Press, 1967.

Sanders, S., Laurendeau, M., & Bergeron, J. Aging and the concept of space: The conservation of surfaces. *Journal of Gerontology,* 1966, **21**, 281–286.

Sarbin, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 1. Cambridge, Mass.: Addison-Wesley, 1954.

Schaie, K. W. A test of behavioral rigidity. *Journal of Abnormal and Social Psychology,* 1955, **51**, 604–610.

Schaie, K. W. Rigidity-flexibility and intelligence: A cross-sectional study of the adult life span from 20 to 70 years. *Psychological Monographs,* 1958, **72** (9, Whole No. 462).

Schaie, K. W. A field-theory approach to age changes in cognitive behavior. *Vita Humana,* 1962, **5**, 129–141.

Schaie, K. W., & Strother, C. R. Limits of optimal functioning in superior old adults. *Interdisciplinary Topics in Gerontology,* 1968, **1**, 132–150.

Sherwood, S. Gerontology and the sociology of food and eating. *Aging and Human Development,* 1970, **1**, 61–85.

Shneidman, E. S., & Farberow, N. L. (Eds.) *Clues to suicide.* New York: McGraw-Hill, 1957.

Sinclair-de-Zwart, H., Ajuriaguerra, J. de, Boehme, M., & Tissot, R. Quelques aspects de la désintégration des notions de temps à travers des épreuves morphosyntaxiques de langage et à travers des épreuves opératiores chez les viellards atteints de démence dégénérative. *Bulletin de Psychologie,* 1966, **19**, 745–751.

Smedslund, J. Les origenes sociales de la décentration. In F. Bresson & M. de Montmollin (Eds.), *Psychologie et épistémologie génétiques.* Paris: Dunod, 1966.

Smith, M. E. A study of some factors influencing the development of the sentence in preschool children. *Journal of Genetic Psychology,* 1935, **46**, 182–212.

Soskin, W. F., & John, V. P. The study of spontaneous talk. In R. Barker (Ed.), *The stream of behavior.* New York: Appleton-Century-Crofts, 1963.

Stuart, R. B. Decentration in the development of children's concepts of moral and causal judgment. *Journal of Genetic Psychology,* 1967, **111**, 59–68.

Sullivan, E. V., & Hunt, D. E. Interpersonal and objective decentering as a function of age and social class. *Journal of Genetic Psychology,* 1967, **110**, 199–210.

Sullivan, H. S. *The interpersonal theory of psychiatry.* New York: Norton, 1953.

Tallmer, M., & Kutner, B. Disengagement and the stresses of aging. *Journal of Gerontology,* 1969, **24**, 70–75.

Turiel, E. Developmental processes in the child's moral thinking. In P. H. Mussen, J. Langer, & M. Covington (Eds.), *Trends and issues in developmental psychology.* New York: Holt, Rinehart & Winston, 1969.

Vygotsky, L. *Thought and language.* Cambridge, Mass.: M. I. T. Press, 1962.

Wapner, S. Organismic-developmental theory: Some applications to cognition. In P. H. Mussen, J. Langer, & M. Covington (Eds.), *Trends and issues in developmental psychology.* New York: Holt, Rinehart & Winston, 1969.

Weinberg, N. H. Relativism, self-centering, and conceptual level. *Child Development,* 1963, **34**, 443–450.

Weir, R. *Language in the crib.* The Hague: Morton, 1962.

Werner, H., & Kaplan, B. *Symbol formation.* New York: Wiley, 1963.

Zinberg, N. E. The relationship of regressive phenomena to the aging process. In N. E. Zinberg & I. J. Kaufman (Eds.), *Normal psychology of the aging process.* New York: International Universities Press, 1963.

(Received November 1, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 2, 93–102

# LOOKING FOR CONFIGURALITY IN CLINICAL JUDGMENT[1]

NORMAN H. ANDERSON[2]

*University of California, San Diego*

Despite a general belief that clinical judgment is complex and configural, previous surveys have found little supportive evidence. Failure to uncover configural processes may reflect inadequate methodology, namely, the use of additive or linear models analyzed by regression-correlation techniques. Functional measurement procedure solves many of these methodological problems, especially the basic problem of measuring the subjective values of the stimuli at the level of the individual judge. In an experimental illustration, judgments of disturbed behavior followed an averaging model; stimulus items that were more extreme in value also had greater weight reflecting their greater diagnosticity. The averaging process, which has received extensive empirical support, is configural because the effective weight of each stimulus depends on the whole set. Other known configuralities, inconsistency discounting, redundancy discounting, and attentional discounting, can also be interpreted in terms of the weight parameter of information integration theory.

Clinical judgment can usefully be considered as information integration. The clinician is faced with a variegated set of informational stimuli that bear on some judgment or decision. His task is to integrate these separate pieces of information into some unitary judgment. One of the most popular hypotheses has been that the clinician's integration rule is "configural," depending in some unique way on the interrelation and configuration of the array of information (see e.g., Goldberg, 1968; Hammond, Hursch, & Todd, 1964; Meehl, 1954; Sawyer, 1966; Wiggins & Hoffman, 1968). Asch (1946) has argued similarly for the configural view in personality impression formation. But as attractive as the configural hypothesis is, little solid evidence has been obtained to support it.

Much of the work on clinical judgment has been summarized by Goldberg (1968) whose main conclusion was that a linear additive model appeared to predict the clinician's judgments quite well. On the face of it, in fact, there was very little evidence for nonlinear or additive processes. Goldberg concluded that the most attractive interpretation was that "the power of the linear regression model is so great that it serves to obscure the real configural processes in judgment [p. 488]." As this quote indicates, Goldberg was sympathetic to the idea of configural judgment, despite the lack of evidence. Indeed, there was even less evidence than Goldberg thought. Of the three or four studies he cited for the configural view, only one, on stockbroker's decisions (Slovic, 1969), presents satisfactory evidence.

*This article shows that available methods, properly applied, have ample power to detect configurality in judgment. It suggests also that configural judgment is quite common, and furthermore that a good deal is already known about what kinds of configurality there are.*

One main difficulty with previous work has been its reliance on standard multiple regression methodology. Those procedures had worked well in practical problems of clinical prediction. It was natural, therefore, to transfer them directly to the study of the judgmental process itself. But that transfer is accompanied by a shift in emphasis and application that often went unrecognized. The failure to detect configurality seems to have resulted in part from failure to use the full capabilities of the multiple regression methodology, in part from the inherent limitations of that methodology. No less seriously, the configuralities or nonlinearities that have been claimed may represent a measurement artifact without psychological meaning.

---

[1] This work was supported by funds from the Research Committee, University of California, Los Angeles, and by National Science Foundation Grant GB-21028. I owe special thanks to Vera Kanareff Presbie for the list of Bizarre Behavior sentences. I also wish to thank Sue Smith for her assistance in running this experiment.

[2] Requests for reprints should be sent to Norman H. Anderson, Department of Psychology, University of California, San Diego, La Jolla, California, 92037.

Many of the points to be made can be most conveniently illustrated in actual data analysis. This is done below. However, a few preliminary remarks on theoretical and methodological background are necessary.

The theoretical background is the writer's theory of information integration. The methodological techniques are those used in functional measurement theory. This material has been discussed elsewhere (Anderson, 1970, 1971a, [1972][3]) and need not be repeated here. However, the averaging hypothesis for information integration has special relevance to the configural view and requires a brief look.

A considerable accumulation of evidence indicates that people integrate information by averaging it. For two stimuli, 1 and 2, the averaging model can be written as:

$$R_{12} = \frac{w_1 s_1 + w_2 s_2}{w_1 + w_2} = \frac{w_1}{w_1 + w_2} s_1 + \frac{w_2}{w_1 + w_2} s_2. \qquad [1]$$

Here $R_{12}$ is the response to the pair of stimuli; $s_1$ and $w_1$ are the scale value and weight of Stimulus 1; and $s_2$ and $w_2$ are the scale value and weight for Stimulus 2. The coefficient of $s_1$ in Equation 1 is $w_1/(w_1 + w_2)$, which is its *relative* weight or importance in the judgment. The relative weights sum to one, the condition for an average. Similar expressions would hold for sets with more than two pieces of information.

The averaging model is usually tested with stimulus combinations constructed from a factorial design. When the levels of each factor have equal weight, the analysis is especially simple. A graphic test of parallelism is then applicable, and ordinary analysis of variance may be applied directly to test the model.

However, equal weighting appears to be the exception rather than the rule. In most judgmental situations, the stimuli along any one dimension will differ not only in scale value but also in weight or importance. The averaging model then becomes nonlinear and considerably more difficult to analyze. One purpose of this report is to illustrate methods for handling the nonlinear case.

In the experiment considered here, the subjects judged severity of disturbance in patients described by their ward behavior. For these stimuli, differential weighting is to be expected. The sentence, *He mumbles to himself when he is denied something that he has requested*, not only implies a less severe degree of disturbance (scale value), but it is also less diagnostic (weight) than the sentence. *He said to the examining physician, "Don't bother with me, I'm no good."*

The averaging model has special psychological interest because it is configural without being unpredictable. Since the relative weights must add to one, the role of any one stimulus depends on the others with which it is combined. A stimulus of moderate absolute importance can dominate the judgment in a group of weak stimuli, but be lost to sight in group of strong stimuli. The response thus depends uniquely on the stimulus configuration. Nevertheless, the response is completely predictable because the same absolute weight applies in all combinations.

## EXPERIMENTAL TASK AND DESIGN

The subjects received descriptions of hypothetical patients hospitalized for mental illness. Each patient was described by four sentences characterizing his ward behavior. The subjects judged the degree of disturbance of each patient on the basis of the given information. The experimental design allows an analysis at the level of each individual subject. It is somewhat complicated by various counterbalancing requirements, but these are unnecessary for an understanding of the main results.

*Stimuli.* The ward behavior sentences were from a large list developed by Kanareff[4] whose normative data allowed a preliminary categorization of sentence scale value. Four sentences were chosen from each of five scale ranges, A, B, C, D, and E, of increasing severity of disturbance. Illustrative sentences are:

---

[3] N. H. Anderson. *Information integration theory: A brief survey.* (Tech. Rep. No. 24) San Diego, Calif.: University of California, Center for Human Information Processing, 1972.

[4] V. Kanareff. Small group decision-making as a function of the distribution and discrepancy of problem-relevant information. Unpublished doctoral dissertation, University of Wisconsin, 1957.

A. During the interviews, he would occasionally be absorbed in the discussion;

B. He mumbles to himself when he is denied something that he has requested;

C. He told the interviewer, "I live for myself and nothing stimulates or interests me";

D. He said to the examining physician, "Don't bother with me, I'm no good";

E. He sometimes walks around shouting "Mother, Mother."

*Design.* Each patient description contained four ward behavior sentences. Ten sentence pairs were used, two from each of the five stimulus ranges, so that the two sentences of a pair had approximately equal value. Each pair was presented with every other pair for a total of 45 patient descriptions.

Each subject judged the 45 descriptions once each day for 5 consecutive days. Order of presentation was balanced for polarity of preceding description. This balanced possible order effects, and also allowed an assessment of carry-over effects from one description to the next. Over Days 3–5, there was a small but significant *assimilation* effect; the mean response was .27 greater when the preceding stimulus was severe than when it was mild. This result is contrary to adaptation-level theory (Helson, 1964) but in agreement with recent work (Anderson, 1971b). The design also controlled the order of presentation of the four sentences within descriptions. There was some evidence of order of presentation effects for Days 1 and 2 so only Days 3–5 were analyzed further.

*Procedure.* Subjects were told that the sentences were observations of ward behavior of patients at Mendota State Hospital who had been hospitalized for disturbed behavior, that some patients were seriously ill whereas others were essentially normal and would shortly be discharged. Each set of four sentences was to be considered a typical cross section of some patient's behavior, the patients themselves being hypothetical although the single sentences were real. Six warm-up descriptions and the 45 experimental descriptions were then presented each day.

Each sentence was typed on a small card, and the set of four placed face down in front of the subject. The subject turned up each successive sentence, read it, turned it down. After reading all four sentences, he turned them all up together, reviewed them for about 20 seconds, and then rated the patient on a 20-point numerical scale labeled "normal" to "severely disturbed" at the two ends. Each patient description required about 45 seconds. The subjects were six advanced undergraduate majors in psychology. Each received $7.50 for his services.

## MODEL ANALYSES

### *Adding Model*

Figure 1 plots mean response for the 15 distinct combinations of stimulus value. The pair-combination procedure produces a triangular rather than a factorial design. Nevertheless, a simple adding model predicts that the data points should plot as parallel curves. Instead, the data show a moderate convergence to the right. This nonparallelism is contrary to either an adding model, or an averaging model with equal weighting of each stimulus.

The statistical analysis of the adding model was performed separately for each subject. A least squares procedure for triangular designs (Anderson [5]; Bechtel, 1967) was used to estimate scale values of the 10 sentence-pairs for each subject. These values were used to predict the response to all 45 descriptions. Deviations from predictions on 35 degrees of freedom were then tested against the within-cell error on 90 degrees of freedom. Of the six subjects, five show significant discrepancies from the model in Table 1. Thus, the adding model does not account for the data, and the same holds for the averaging model with equal weighting.

Three methodological points in this analysis of the adding model require brief discussion.

TABLE 1

TEST OF DEVIATIONS FROM PREDICTION, ADDING AND AVERAGING MODELS

| Subject | Adding model | | Averaging model |
|---|---|---|---|
| | $r$[a] | $F$[b] | $F$[c] |
| LK | .97 | 2.70* | .82 |
| PG | .94 | 1.99* | 1.11 |
| CT | .95 | 1.56* | .80 |
| GB | .97 | 2.40* | 1.52 |
| AS | .96 | 1.14 | 1.16 |
| MG | .97 | 1.69* | 1.66* |

[a] Correlation between predicted and observed, adding model.
[b] Tests deviations from adding model, $df = 35/90$.
[c] Tests deviations from averaging model, $df = 33/90$.
* $p < .05$; critical values at .01 and .05 levels are 1.56 and 1.87, respectively.

---

[5] N. H. Anderson. Application of an additive model to impression formation. Paper presented at the third annual meeting of the Psychonomic Society, St. Louis, August 1962.

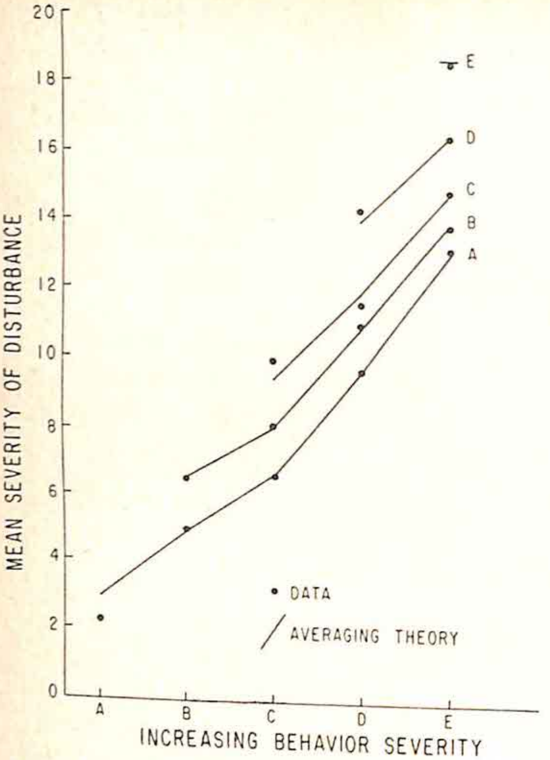


FIG. 1. Mean judgment of severity of disturbance indicated by ward behavior. A, B, C, D, and E represent sentence pairs of increasing severity, and each data point represents the combination of two such sentence pairs. Filled circles are observed results; lines are predictions from averaging theory.

*Correlation coefficients.* It needs to be emphasized that the correlation coefficients of Table 1 are entirely inappropriate for testing the model.[6] They measure the wrong thing, namely, the degree of agreement between model and data. Although the model–data correlations are quite high, averaging .96, the adding model is not correct because the discrepancies are significant. The relevant measure is of degree of disagreement between model and data; the analysis of variance has ample power to detect that.

This basic methodological point has not always been appreciated. Much of the evidence that has been considered to show that human judgment follows a linear additive model rests on such inappropriate statistics. More appropriate design and analysis might well have revealed configural processes. This paper provides one such example.

Correlations may be entirely appropriate for certain purposes, especially in applied prediction situations. In experimental studies of judgment, however, high correlations are typically guaranteed by the choice of experimental stimuli (Anderson, 1962a). At the same time, discrepancies from prediction need not be fatal since they may only reflect nonessential failure of some simplifying assumption. Each case needs separate examination, however. Even relatively small discrepancies may reflect a basic flaw in the model, as illustrated here for the adding model.

*Multiple regression model.* The failure of the adding model implies that a linear regression model must also fail. Both models have the same mathematical form, but differ in their treatment of the stimulus scale values. In the present analysis of variance, the scale values are estimated for each subject from his data. Standard multiple regression, in contrast, requires previously known scale values.

Doubts sometimes arise whether it is fair to estimate parameter values from the very data to be predicted. With linear models, such tests are completely legitimate since each parameter estimate uses up one degree of freedom. Of more immediate importance is that such scale values are in a certain sense the best possible. The regression analysis could not, therefore, be more accurate. Consequently, the failure of the adding model under the analysis of variance test implies a corresponding failure of the linear regression model.

This last point brings up two important limitations of regression analysis, both of which stem from its dependence on prior estimates of the stimulus values. This dependence might seem a virtue, but in fact it is a defect in two respects. In many cases, these stimulus values will be subject to random error. That needs to be taken into account, and it is possible to do so (e.g., Williams, 1959). The usual applications do not do that, however, and hence are not always correct.

Also, in many cases, the stimulus values will be subject to constant errors. For example, the subject's set may shift between the scaling task and the integration task. Any such constant error appears as a discrepancy from the regression model, causing it to fail. But that means that significant discrepancies from a regression model are ambiguous. They may reflect a shortcoming in the model itself, or they may merely reflect incorrect stimulus values.

---

[6] The situation is even worse. A striking paper by Birnbaum (in press) shows that correlations can easily be higher for an incorrect model than for a correct model.

Functional measurement avoids this shortcoming of regression analysis. Scaling is done within the integration task itself. Significant deviations from the model will not reflect incorrect stimulus values per se.

*Response scale.* The above analysis requires that the response measure is an equal-interval scale. If the model fits the data, no problem may arise. In that case, the test of fit gives joint support to the model and the response scale. Success of the model simultaneously constitutes a functional validation of the interval property of the response scale (Anderson, 1970; see Footnote 3).

If the model fails the test, however, then the response scale may come under suspicion. The discrepancy could reflect a faulty response scale even though the model itself was correct. Functional measurement attacks this problem with a monotone response scaling procedure (Anderson, 1962b; see also Bogartz & Wackwitz, 1971). The essential idea is simple. If the response scale is at least ordinal, then a monotone transformation will yield an interval scale.

A response transformation was made in the original report of the present data (Anderson, see Footnote 5), but four of the six subjects still had significant discrepancies. Only the quadratic expansion was used, and a higher-order polynomial might still have saved the adding model. However, the cumulated experimental evidence since then argues fairly conclusively against the adding model; it cannot account for the cross-over interactions obtained in repeated experiments (see Anderson, 1971a).

### *Averaging Model*

The predictions of the averaging model with differential weighting are shown by the solid curves of Figure 1. These theoretical curves reproduce the rightward convergence pattern of the data. The fit is quite close; the mean magnitude of discrepancy between data and theory is only .2. In the rightmost column of Table 1, only one of the six subjects shows a significant discrepancy. In contrast to the adding model, the averaging model does rather well.

Functional measurement procedures were applied to scale the stimulus values and weights used in the above predictions. The unit stimulus was actually a matched pair of sentences, and functional scales were obtained separately for each subject, thus allowing for individual differences. Since weight could be expected to vary directly with scale value in the present case, $w$ was taken as a quadratic function of $s$:

$$w = 1 + as + bs^2. \qquad [2]$$

Equation 2 represents an approximate method that may be widely useful. Its advantage is that only two additional parameters, $a$ and $b$, are required to estimate the 10 weights. The estimation was done using Chandler's[7] computer routine. The initial scale values were those estimated from the adding model; the initial values of $a$ and $b$ were zero. About 15 seconds were required to estimate 10 scale values and two weight parameters for each subject.

The weights turned out to be nearly linear functions of the scale values; the estimates of $b$ were near zero and added little to the fit. Indeed, the $F$ ratios for discrepancy were slightly larger for half the subjects when $b$ was estimated than when $b$ was set equal to zero. To emphasize this point, the theoretical curves of Figure 1 were obtained by estimating only $a$ with $b$ set equal to zero. The mean value of $a$ was .023. If the weight of the least severe behavior is taken as 1, the weight of the most severe behavior is 1.88, a ratio of not quite 2 to 1. In this judgment task, the differential weighting is relatively mild.

Three comments on this analysis should be added. First, no allowance was made for an initial impression, $I_0$. In the present task, its effect was expected to be small, but that would not always be the case (e.g., Leon, Oden, & Anderson, in press). Second, Equation 1 is nonlinear when the weights are estimated. The $F$ test is then only approximate since the estimates are not completely orthogonal. However, that would only become important when the number of weight parameters to be estimated was relatively large, and in any case would presumably yield a conservative $F$ test. Third, the justification for the averaging model rests not simply on its ability to fit the present data but also on its success in related work, which led to the present reanalysis. The adding model might be modified in various ways to give a better fit, but differential weighting by itself is not one of them. Even with differential weighting, the adding model still predicts parallelism (Anderson, 1971a, p. 180). Rescaling the response measure, as noted above, might perhaps preserve the noninteractive adding model for these particular data. However, the adding model has failed repeatedly in critical tests with similar stimuli (Anderson, 1971a, p. 181).

---

[7] J. P. Chandler. *Subroutine Stepit.* (Program QCPE 66) Quantum Chemistry Program Exchange. Bloomington, Ind.: Indiana University, 1965.

### *Additional Technical Comments*

*Regression analysis.* Although regression-correlation technique has been exceedingly popular in the study of judgment, it has numerous severe limitations. Most of these have been discussed elsewhere (e.g., Anderson, 1962a, 1969, 1971a, pp. 193–194; Birnbaum, in press; Bogartz & Wackwitz, 1971; Darlington, 1968; Hoffman, Slovic, & Rorer, 1968; Slovic & Lichtenstein, 1971) and are not repeated here. However, one implication of differential weighting deserves specific note.

In the present analysis, items that were more extreme in value were also more extreme in weight. Such differential weighting must be quite common in clinical judgment, with the Minnesota Multiphasic Personality Inventory (MMPI) scales, for instance. Ordinary multiple regression cannot handle differential weighting because it assumes a constant beta weight within each stimulus dimension. Not even the extension of functional measurement to regression analysis by Bogartz and Wackwitz (1971) can handle this problem. The weight cannot be transformed into the value dimension because weight and value are strictly distinct in the averaging model.

It should perhaps be repeated that regression analysis is discussed here with specific reference to its usual applications in process models of judgment. Regression analysis has many useful properties, more flexibility than is always utilized, and it is the appropriate tool in many situations.

*Nonlinearity in judgment.* A good deal of work in clinical judgment rests on the implicit assumption that "nonlinearity" in a regression analysis has direct psychological significance. That is not true in general; any such inference requires explicit justification. Nonlinearity must be measured in terms of the subjective scale. A process that is linear in the subjective values will necessarily appear to be nonlinear unless equal-interval scales are used for stimulus and response. Unless it is known that the measurements have been made on subjectively equal-interval scales, any observed nonlinearity may be meaningless.

An interesting illustration of this problem arises in Goldberg's (1971) comparison of one linear and four nonlinear models for representing clinician's predictions of psychosis from MMPI scales. As Birnbaum (in press) points out, the five models differ only by arbitrary scaling of the stimulus or response variables. Indeed, they were all analyzed by reduction to a linear form. To single out any one as linear in the psychological sense requires singling out one set of scales as being the true interval scales. Since there is no great reason to believe that the MMPI scales are equal interval, Goldberg's linear–nonlinear distinction is tenuous. To be fair to Goldberg, it should be stated again that different usage may be allowable for practical prediction than for study of psychological processes. For prediction, it may be useful to arbitrarily define some convenient set of scales as being equal interval, and interpret nonlinearity relative to them. But an understanding of the judgmental process clearly must be in terms of the subjective values of the judge. Only in that metric does nonlinearity have any real meaning.

Without a theory of measurement, the problem is not resolvable. And it should be noted that not even Thurstonian paired comparisons will provide interval scales for the typical stimuli of clinical judgment. In practice, such scales require pooled group data (Bock & Jones, 1968, p. 4; Leon et al., in press). Individuals would seldom be long in doubt as to the rank order of two clinical stimuli so their choice proportions would be largely 0 or 1. If all individuals had the same rank order, then the group proportions would also be 0 or 1. To get usable choice proportions requires pooling individuals who differ in their rank ordering of the stimuli. The paired-comparison scales are constructed out of real individual differences. They cannot, therefore, adequately represent the separate individuals.

*Configurality*. The multiple regression approach to configurality has typically proceeded by adding powers and cross-products of the predictor variables to the linear regression model. It has been assumed (e.g., Meehl, 1954; Wiggins & Hoffman, 1968) that significant beta weights for the cross-product terms would imply configurality in the judgmental process. This assumption is invalid (Bogartz & Wackwitz, 1971). The cross-product terms in multiple regression are analogous to interaction terms in analysis of variance, which are discussed below. Moreover, a cross-product term, even if real, need not imply configural process. For example, the multiplying model for estimated area of rectangles (Anderson & Weiss, 1971) could result from an additive integration.

*Analysis of variance*. The analysis of variance has proved far better than could have been expected from a general purpose statistical technique. Besides work by the writer, mention should be made of important applications by Slovic and his colleagues at the Oregon Research Institute (e.g., Hoffman et al., 1968; Slovic, 1969; Slovic & Lichtenstein, 1971). The fundamental advantage of analysis of variance is that it does not require a prior stimulus metric, but instead allows direct estimation of subjective values. That is one reason why it is so useful in functional measurement.

Four common misconceptions about the role of analysis of variance in judgment research should also be noted. First, significant $F$ ratios for interaction terms do not, unfortunately, necessarily imply patterning, configural response, or stimulus interaction in any psychological sense (see e.g., Anderson, 1961, Figure 1; Bogartz & Wackwitz, 1971). The term "interaction" in statistics is as regrettable as the term "significant," since neither has its common English meaning.

Second, the stimulus variables need not be categorical, but can be continuous, numerical dimensions. If a stimulus metric is available, then the ordinary linear orthogonal polynomials are equivalent to the linear multiple regression. Further, the complete orthogonal polynomial analysis is equivalent to a complete regression analysis, all powers and cross-products included. Of course, such nonlinear terms have meaning only in terms of the stimulus metric. They may be purely nominal, without psychological meaning, unless the subjective stimulus values are used.

Third, a complete factorial design is not necessary (e.g., Winer, 1971, Chapter 5.23). Balanced designs have important statistical advantages, but psychological considerations have first priority in experimental design. It is entirely sensible to argue, as Brunswik (1956) and others have done, that stimulus combinations that do not make sense should be omitted. But that is an argument against routine, unthinking use of factorial design, and has essentially nothing to do with analysis of variance.

Fourth, analysis of variance imposes no requirement to use uncorrelated cues in the design. Indeed, the objective or actuarial correlation of the cues is irrelevant since the judgment depends on how the subject uses the cues. If any subjective correlation, reflecting inconsistency or redundancy, for instance, manifests itself as a significant interaction, then analysis of variance has done its duty if it gives a warning signal. If analysis of variance is to be criticized in this respect, it is when it fails to reveal such subjective interactions. Examples of this are known (Anderson, 1966; Anderson, Lindner, & Lopes, in press), though the matter has not been studied systematically.

## DISCUSSION

### *Configurality in Judgment*

The general view that emerges from experimental studies of information integration is that configural processes pervade judgment. The most important single configural process is information averaging. Averaging processes have particular interest because, as noted in the introduction, they are configural without being unpredictable. Although the relative weight of each stimulus will depend on what other stimuli it is combined with, these relative weights can all be calculated directly from the absolute weights of the separate stimuli.

That averaging is configural is obscured in certain applications. When the stimuli along each dimension are equally weighted, then the averaging model has a simple linear form. However, when differential weighting prevails, as in the present experiment, then nonlinearity becomes clear (e.g., Lampel & Anderson, 1968; Leon et al., in press; Oden & Anderson, 1971). Einhorn's (1970, 1971) conjunctive and disjunctive models may also be interpretable as differential weighting (Anderson, 1971a).

More complex kinds of configurality have also been indicated by the experimental evidence. These include inconsistency discounting (e.g., Anderson & Jacobson, 1965; Bugental, Kaswan, & Love, 1970; Hendrick, 1972; Schümer & Cohen, 1968), redundancy discounting (e.g., Kaplan, 1971a; Schmidt, 1969; Wyer, 1968), and cognitive simplification of complex stimulus fields (e.g., Leon et al., in press; Shanteau & Anderson, 1972; Shepard, 1964; Slovic & Lichtenstein, 1971). Clear-cut results on these more complex configuralities tend to be difficult, however, and there is a need for simple, general experimental methods for their analysis.

If it would work, the most useful tool for studying configural processes would be the method of component judgments. If the weight and value parameters depend uniquely on the stimulus array, they might be determined most directly by asking the subject to judge them in context. What work there is along these lines presents a mixed picture. Judgments of component scale value are readily obtained but are subject to strong context effects (e.g., Anderson, 1966; see Footnote 3; Kaplan, 1971b). These context effects may be linear, however, in which case they could still be quite useful.

Judgments of components weight may be more difficult. Some results have been promising (Anderson & Alexander, 1971), but other work has been less optimistic (see Anderson et al., in press; Slovic, 1969; Slovic & Lichtenstein, 1971). Various technical difficulties, separating weight from scale value, for example, have complicated the interpretation of existing studies and held back the design of further work. The potential usefulness of component weight judgments can hardly be overrated, however, and it is to be hoped that further study will resolve the difficulties.

It is interesting to note the importance of the weight parameter in configural judgment. Its role in averaging theory has already been discussed. Discounting itself reflects reduced weighting, from inconsistency or from redundancy in the two cases. Cognitive simplification can be interpreted as attentional discounting; simplifying the stimulus field by ignoring certain stimuli corresponds to setting their weights equal to zero. The weighting process may thus provide a unifying concept for configural judgment.

### *Clinical Judgment*

Most of the evidence for configural judgment has been obtained in laboratory experiments on person perception. Since person perception is often considered apart from clinical judgment, some comments on the relevance of the present methodology are in order.

Perhaps the major difference between laboratory and field situations is the degree of experience and training of the judges. The differences between naive and experienced judges will appear in two theoretically distinct ways: valuation and integration (Anderson, 1971a; see Footnote 3). Experience will certainly affect the valuation process; an MMPI scale might be virtually meaningless to an untrained judge, for example. Most practical prediction tasks differ from the person perception studies in the obvious necessity for special training in valuation operations.

Nevertheless, the integration rules themselves may well be basically similar for naive and experienced judges. The averaging effect has been found in so many different stimulus situations that it is a candidate for a general principle of judgment. The other configural processes mentioned above must also be expected to occur in experienced judges. Indeed, experienced judges would probably be more sensitive than naive judges to redundancy and, in most tasks, to inconsistency. On the other hand, experienced judges would probably make less use of simplification strategies.

The distinction between valuation and integration parallels, to a certain degree, the difference between parameters and functional rule. In the functional measurement approach, the integration rule is used as the primary basis for measuring the stimulus parameters, scale value and weight. This allows numerical scaling of qualitative stimuli, at the level of the individual judge, as has been demonstrated here. It can even measure internal states, as in Kaplan's[8] notable work on response dispositions (see also Kaplan & Anderson, in press).

The present approach places first emphasis on the judgmental process itself. Measurement becomes an organic part of the substantive investigation, facilitating a focused search for configural processes of judgment. The stony field of statistical versus clinical prediction is bypassed. This agrees with Goldberg's (1968) orientation, as well as with numerous other researchers on judgment (e.g., Cronbach, 1955; Holt, 1970, p. 347; Sarbin, Taft, & Bailey, 1960; Slovic & Lichtenstein, 1971; Tagiuri, 1969, pp. 414 ff). The methods illustrated here should provide useful leverage on the problem.

---

[8] M. F. Kaplan. *How response dispositions integrate with stimulus information.* (Tech. Rep. No. 19) San Diego, Calif.: University of California, Center for Human Information Processing, 1971.

## REFERENCES

Anderson, N. H. Scales and statistics: Parametric and nonparametric. *Psychological Bulletin,* 1961, **58,** 305–316.

Anderson, N. H. Application of an additive model to impression formation. *Science,* 1962, **138,** 817–818. (a)

Anderson, N. H. On the quantification of Miller's conflict theory. *Psychological Review,* 1962, **69,** 400–414. (b)

Anderson, N. H. Component ratings in impression formation. *Psychonomic Science,* 1966, **6,** 279–280.

Anderson, N. H. Comment on "An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment." *Psychological Bulletin,* 1969, **72,** 63–65.

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review,* 1970, **77,** 153–170.

Anderson, N. H. Integration theory and attitude change. *Psychological Review,* 1971, **78,** 171–206. (a)

Anderson, N. H. Test of adaptation-level theory as an explanation of a recency effect in psychophysical integration. *Journal of Experimental Psychology,* 1971, **87,** 57–63. (b)

Anderson, N. H., & Alexander, G. R. Choice test of the averaging hypothesis for information integration. *Cognitive Psychology,* 1971, **2,** 313–324.

Anderson, N. H., & Jacobson, A. Effect of stimulus inconsistency and discounting instructions in personality impression formation. *Journal of Personality and Social Psychology,* 1965, **2,** 531–539.

Anderson, N. H., Lindner, R. C., & Lopes, L. L. Integration theory applied to judgments of groups. *Journal of Personality and Social Psychology,* in press.

Anderson, N. H., & Weiss, D. J. Test of a multiplying model for estimated area of rectangles. *American Journal of Psychology,* 1971, **84,** 543–548.

Asch, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology,* 1946, **41,** 258–290.

Bechtel, G. G. The analysis of variance and pairwise scaling. *Psychometrika,* 1967, **32,** 47–65.

Birnbaum, M. H. The Devil rides again: Correlation as an index of fit. *Psychological Bulletin,* in press.

Bock, R. D., & Jones, L. V. *The measurement and prediction of judgment and choice.* San Francisco: Holden-Day, 1968.

Bogartz, R. S., & Wackwitz, J. H. Polynomial response scaling and functional measurement. *Journal of Mathematical Psychology,* 1971, **8,** 418–443.

Brunswik, E. *Perception and the representative design of experiments.* Berkeley, Calif.: University of California Press, 1956.

Bugental, D. E., Kaswan, J. W., & Love, L. R. Perception of contradictory meanings conveyed by verbal and nonverbal channels. *Journal of Personality and Social Psychology,* 1970, **16,** 647–655.

Cronbch, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin,* 1955, **52,** 177–193.

Darlington, R. B. Multiple regression in psychological research and practice. *Psychological Bulletin,* 1968, **69,** 161–182.

Einhorn, H. J. Use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin,* 1970, **73,** 221–230.

Einhorn, H. J. Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance,* 1971, **6,** 1–27.

Goldberg, L. R. Simple models or simple processes? Some research on clinical judgments. *American Psychologist,* 1968, **23,** 483–496.

Goldberg, L. R. Five models of clinical judgment: An empirical comparison between linear and nonlinear representations of the human inference process. *Organizational Behavior and Human Performance,* 1971, **6,** 458–479.

Hammond, K. R., Hursch, C. J., & Todd, F. J. Analyzing the components of clinical inference. *Psychological Review,* 1964, **71,** 438–456.

Helson, H. *Adaptation-level theory.* New York: Harper & Row, 1964.

Hendrick, C. Effects of salience of stimulus inconsistency on impression formation. *Journal of Personality and Social Psychology,* 1972, **22,** 219–222.

Hoffman, P. J., Slovic, P., & Rorer, L. G. An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. *Psychological Bulletin,* 1968, **69,** 338–349.

HOLT, R. R. Yet another look at clinical and statistical prediction: Or, is clinical psychology worthwhile? *American Psychologist,* 1970, **25**, 337–349.

KAPLAN, M. F. The determination of trait redundancy in impression formation. *Psychonomic Science,* 1971, **23**, 280–282. (a)

KAPLAN, M. F. Context effects in impression formation: The weighted average versus the meaning-change formulation. *Journal of Personality and Social Psychology,* 1971, **19**, 92–99. (b)

KAPLAN, M. F., & ANDERSON, N. H. Information integration theory and reinforcement theory as approaches to interpersonal attraction. *Journal of Personality and Social Psychology,* in press.

LAMPEL, A. K., & ANDERSON, N. H. Combining visual and verbal information in an impression-formation task. *Journal of Personality and Social Psychology,* 1968, **9**, 1–6.

LEON, M., ODEN, G. C., & ANDERSON, N. H. Functional measurement of social values. *Journal of Personality and Social Psychology,* in press.

MEEHL, P. E. *Clinical vs. statistical prediction.* Minneapolis: University of Minnesota Press, 1954.

ODEN, G. C., & ANDERSON, N. H. Differential weighting in integration theory. *Journal of Experimental Psychology,* 1971, **89**, 152–161.

SARBIN, T. R., TAFT, R., & BAILEY, D. E. *Clinical inference and cognitive theory.* New York: Holt, Rinehart & Winston, 1960.

SAWYER, J. Measurement *and* prediction, clinical *and* statistical. *Psychological Bulletin,* 1966, **66**, 178–200.

SCHMIDT, C. F. Personality impression formation as a function of relatedness of information and length of set. *Journal of Personality and Social Psychology,* 1969, **12**, 6–11.

SCHÜMER, R., & COHEN, R. Eine Untersuchung zur sozialen Urteilsbildung: II. Bemerkungen zu verschiedenen konkurrierenden Modellen der Urteilsbildung. *Archiv für die gesamte Psychologie,* 1968, **120**, 180–202.

SHANTEAU, J. C., & ANDERSON, N. H. Integration theory applied to judgments of the value of information. *Journal of Experimental Psychology,* 1972, **92**, 266–273.

SHEPARD, R. N. On subjectively optimum selections among multi-attribute alternatives. In M. W. Shelley & G. L. Bryan (Eds.), *Human judgments and optimality.* New York: Wiley, 1964.

SLOVIC, P. Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology,* 1969, **53**, 255–263.

SLOVIC, P., & LICHTENSTEIN, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance,* 1971, **6**, 649–744.

TAGIURI, R. Person perception. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology.* Vol. 3. (2nd ed.) Reading, Mass.: Addison-Wesley, 1969.

WIGGINS, N. L., & HOFFMAN, P. J. Three models of clinical judgment. *Journal of Abnormal Psychology,* 1968, **73**, 70–77.

WILLIAMS, E. J. *Regression analysis.* New York: Wiley, 1959.

WINER, B. J. *Statistical principles in experimental design.* (2nd ed.) New York: McGraw-Hill, 1971.

WYER, R. S. The effects of information redundancy on evaluations of social stimuli. *Psychonomic Science,* 1968, **13**, 245–246.

(Early publication received April 4, 1972)

Psychological Bulletin
1972, Vol. 78, No. 2, 103–118

# DIFFERENCES BETWEEN BEHAVIOR REINFORCED BY ELECTRICAL STIMULATION OF THE BRAIN AND CONVENTIONALLY REINFORCED BEHAVIOR:

## AN ASSOCIATIVE ANALYSIS[1]

IRMINGARD I. LENZER[2]

*Saint Mary's University, Halifax, Nova Scotia, Canada*

The literature dealing with the differences between behavior maintained by rewarding brain stimulation and behavior maintained by conventional rewards is briefly reviewed. The literature is then interpreted in an associative framework, with the emphasis on specifying the stimuli that are controlling the behavior in each situation. The associative analysis shows that rewarding brain stimulation is different from conventional rewards in some fundamental respects, but that, nevertheless, the behavior maintained by rewarding brain stimulation does, in certain situations, resemble behavior maintained by conventional rewards. Brief critiques of some of the earlier theories are offered from the standpoint of the associative model.

Perhaps the major question in the study of behavior maintained by rewarding electrical stimulation of the brain (ESB) is whether or not ESB reward is fundamentally different from conventional rewards such as food and water. Some studies show that behavior maintained by ESB reward differs from behavior maintained by conventional rewards; several theories—including the drive decay (Deutsch, 1963; Deutsch & Deutsch, 1966; Deutsch & Howarth, 1963; Gallistel, 1964), the aversive aftereffects (Ball, 1967; Ball & Adams, 1965; Miller, 1961a), and the conflict (Kent & Grossman, 1969; Reid, Wasden, & Courtney, 1970; Wasden & Reid, 1968; Wasden, Reid, & Porter, 1965; Wetzel, 1963)—have been proposed to account for these behavioral differences in terms of a difference in the nature of the rewards. Other studies suggest that the behavioral differences can be reduced or eliminated by appropriate manipulations of training and deprivation conditions; this has led to an incentive theory (Trowill, Panksepp, & Gandelman, 1969) based on the idea that ESB reward has the same essential nature as conventional rewards. The purpose of this paper is to reinterpret these studies in an associative framework and to propose a compromise theory that can incorporate and reconcile some of the previous theories.

### BEHAVIORAL DIFFERENCES

The apparent differences between ESB-reinforced and conventionally reinforced behaviors can be divided into six categories.

#### *Lability*

Some studies show that ESB-maintained behavior is labile relative to conventionally reinforced behavior in that it deteriorates more rapidly as a result of the temporary or permanent withdrawal of the primary reward. This lability has been reported in four different situations.

*Resistance to extinction.* Some studies show that behavior maintained by ESB reward on a continuous reinforcement (CRF) schedule deteriorates very rapidly when the reward is withdrawn (Culbertson, Kling, & Berkley, 1966; Gibson, Reid, Sakai, & Porter, 1965; Howarth & Deutsch, 1962; Olds, 1955; Olds & Milner, 1954; Pliskoff & Hawkins, 1963; Seward, Uyeda, & Olds, 1959). However, other studies have shown that this rapid extinction is not an inherent or unalterable property of ESB-maintained behavior. Such factors as food deprivation (Deutsch & Dicara, 1967), conditioned inhibition (Herberg & Watkins, 1966), the spatio-temporal relation between response and reward (Gibson et al., 1965), and the length of delay between the acquisition and the extinction periods (Quartermain & Webster, 1968) have been shown to influence the rate of extinction. Panksepp and Trowill (1967a, 1967b) presented evidence to show that the conventionally reinforced response extinguishes just as rapidly as the ESB-maintained response, provided that both the deprivation level and the spatio-temporal relation between response and reward are the same for the two groups of animals.

---

[1] The preparation of this paper was supported in part by Grant APA 0380 from the National Research Council of Canada.

[2] Requests for reprints should be sent to Irmingard Lenzer, Department of Psychology, Saint Mary's University, Halifax, Nova Scotia, Canada.

*Partial reinforcement performance*. Some studies show that animals responding for ESB reward on a partial reinforcement schedule perform more poorly than animals responding for conventional rewards (Brady & Conrad, 1960; Brodie, Moreno, Malis, & Boren, 1960; Culbertson et al., 1966; Elder, Montgomery, & Rye, 1965; Keesey & Goldstein, 1968; Pliskoff, Wright, & Hawkins, 1965; Sidman, Brady, Boren, Conrad, & Schulman, 1955). However, other studies have shown that the partial reinforcement performance of animals receiving ESB reward can be improved by various experimental manipulations, to a level commensurate with that of animals responding for conventional rewards. Such variables as the brain area (Brodie et al., 1960), the level of food deprivation (Elder et al., 1965), the spatio-temporal relation between response and reward (McIntire & Wright, 1965; Pliskoff et al., 1965), the number of ESB trains per reinforcement (Brown & Trowill, 1970; Pliskoff et al., 1965), and the rate of the shift from the CRF to the partial reinforcement schedule (Elder et al., 1965; Huston, 1968; Pliskoff et al., 1965) have been studied in this connection. Also, certain drugs, such as methamphetamine (Elder et al., 1965) and chlordiazepoxide (Trowill et al., 1969), have been found to improve partial reinforcement performance for ESB reward. Finally, it has been found that in the operant discrimination situation, which is a special subclass of the multiple partial reinforcement situation, the performance for ESB reward is just as good as, if not better than, the performance for conventional rewards (Kling & Berkley, 1968; Lenzer & Frommer, 1968, 1971; Sadowsky, 1969).

*Runway behavior*. Some studies show that animals responding for ESB reward in the runway situation perform more poorly (with slower running speeds and/or slower rates of learning) after long, as opposed to short, intertrial intervals (Gallistel, 1966, 1967, 1969a; Newman, 1961; Panksepp, Gandelman, & Trowill, 1968; Scott, 1967; Seward, Uyeda, & Olds, 1960; Spear, 1962). Spacing of trials does not have such effects on runway behavior when conventional rewards are used (Cotton & Lewis, 1957; Gallistel, 1967; Porter, 1938; Sheffield, 1949; Wilson, Weiss, & Amsel, 1955). However, here again, there are studies that show that this poor performance in the runway situation does not necessarily follow from the use of ESB as the reinforcement. Such factors as the time since the last ESB (Gallistel, 1969a, 1969b), the number of trials since the shift from the short to the long intertrial interval (Panksepp et al., 1968), and the brain area (Scott, 1967) have been found to alter the degree of lability in the runway situation. Also, the drug sodium amytal, a fear-reducing drug (Miller, 1961b), has been found to eliminate the lability in the runway situation (Reid et al., 1970; Wasden & Reid, 1968). Finally, Kent and Grossman (1969) reported that animals that do not require free stimulation (priming) to induce them to perform in the runway do not run more slowly after the long intertrial intervals.

*Overnight decrement*. It has been reported that animals responding for ESB reward in the runway or in the maze situation show an overnight decrement in responding, with slower running time on the first trial of a day than on the last trial of the previous day or on subsequent trials of the same day (Kent & Grossman, 1969; Olds, 1956; Panksepp et al., 1968; Reid et al., 1970; Seward et al., 1960; Spear, 1962; Wasden & Reid, 1968; Wasden et al., 1965). An overnight decrement is sometimes seen with conventional rewards (Logan, 1960, p. 62), but it is not as large as in these studies with ESB reward. However, there are other studies that show that this overnight decrement is not an unalterable property of ESB-maintained behavior. Such factors as brain area (Kent & Grossman, 1969; Wasden et al., 1965), length of intertrial interval (Scott, 1967), and time since last ESB or priming stimulation (Gallistel, 1967; Kent & Grossman, 1969; Panksepp & Trowill, 1968; Wasden et al., 1965; Wetzel, 1963) have been found to affect the amount of overnight decrement. Also, the administration of the drug sodium amytal has been found to reduce or eliminate the overnight decrement, whether the ESB is given to the hypothalamus (Wasden & Reid, 1968) or to the forebrain (Reid et al., 1970). Finally, it has been found that animals trained to respond differentially to stable environmental or internal stimuli do not show an overnight decrement (Kornblith & Olds, 1968; Lenzer & Frommer, 1968).

### *Insatiability*

The second apparent difference is that animals responding for ESB on a CRF schedule may continue to respond at a high and fairly constant rate over many hours at a time, in some cases to the point of physical exhaustion (Olds, 1958; Ray, Hine, & Bivens, 1968), whereas animals responding for conventional rewards show a more rapid tapering off of the response rate.

### *Rate of Acquisition*

There is some evidence that lever-pressing behavior maintained by ESB reward on a CRF schedule (Olds & Milner, 1954) and certain types of discrimination behavior maintained by ESB reward (Lindholm & Keesey, 1970; Sadowsky, 1969) are acquired more rapidly than the corresponding behaviors maintained by conventional rewards. Other evidence, however, suggests that the rapid acquisition of the lever-press response, reinforced by ESB on a CRF schedule, is an artifact of training and deprivation conditions (Gibson et al., 1965; Panksepp & Trowill, 1967b). Moreover, there is some evidence (Kling & Berkley, 1968; Kling & Matsumiya, 1962; Terman & Kling, 1968) that discrimination behavior maintained by ESB reward is not acquired more rapidly than conventionally reinforced discrimination behavior, although this evidence is not as persuasive as that provided by Sadowsky (1969) and Lindholm and Keesey (1970).

### *Dominance*

Under certain conditions an animal will choose to respond for ESB reward instead of conventional reward (Deutsch, Adams, & Metzner, 1964; Spies, 1965) or will respond for ESB reward and fail to respond, in a two-bar competition situation, for conventional reward (Falk, 1961; Mogenson, 1969; Morgan & Mogenson, 1966; Routtenberg & Lindy, 1965), even when the animal is deprived for long periods of time and the conventional reward, food or water, is necessary for survival. However, the ESB-maintained behavior is not always dominant, either in the choice situation or in the two-bar competition situation. By manipulating such variables as brain area (Routtenberg & Lindy, 1965; Spies, 1965), current intensity of the ESB (Deutsch et al., 1964; Falk, 1961; Mogenson, 1969; Morgan & Mogenson, 1966), level of water deprivation (Deutsch et al., 1964), and length of the delay between the priming stimulation and the start of a trial in the choice situation (Deutsch et al., 1964), the ESB-maintained behavior can be made either dominant or subordinate with respect to the conventionally reinforced behavior.

### *Secondary Reinforcement*

There is some evidence (Mogenson, 1965; Seward et al., 1959) that it is more difficult to impart secondary reinforcement properties to a neutral stimulus when that stimulus is paired with ESB reinforcement than when it is paired with conventional reinforcement. However, the important factor in this comparison may be the level of deprivation, rather than the nature of the rewards. The studies by Seward et al. (1959) and Mogenson (1965) involved only nondeprived animals. DiCara,[3] using food-deprived animals, was able to obtain secondary reinforcement effects, provided that the ESB electrode was located in a brain area sensitive to the level of food deprivation. Moreover, some tests for secondary reinforcement in nondeprived animals trained to respond for conventional rewards have also turned out to be negative (Estes, 1949; Schlosberg & Pratt, 1956).

---

[3] L. V. DiCara. Brain stimulation and secondary reinforcement. Unpublished doctoral dissertation, New York University, 1966.

### *Contrast Effects*

Panksepp and Trowill (1969, 1970) presented evidence for negative and positive contrast effects in rats trained to press a lever for hypothalamic ESB reward. Although the negative contrast effect has consistently been obtained with conventional rewards, there is considerable doubt as to the reliability of the positive contrast effect produced by a change in quantity or quality of conventional reinforcement (Black, 1968, 1969; Dunham, 1968; Premack, 1969). The ESB-maintained behavior thus seems to differ from conventionally reinforced behavior in that it is more susceptible to the positive contrast effect.

## ASSOCIATIVE MODEL

The associative model begins with the question, What stimuli are controlling or come to control the ESB-maintained behavior in a given situation? The core idea is that in the CRF situation and in other situations where ESBs follow one another closely, the controlling stimuli (the stimuli that give rise to the operant response) are the internal stimuli, or some subset of the internal stimuli, produced by the ESB. In situations where there is a long interval between one ESB and the next, or between one ESB and the next opportunity for the animal to respond for another ESB, the ESB-produced stimuli, because they decay rapidly with time, cannot serve as controlling stimuli; the animal must learn to respond to other stimuli, if it is to respond at all.

In most situations where conventional rewards are used, and where the rewards follow one another closely in time, the stimuli produced by the reward do not, to any large extent, control the response, though they may control the nonresponse (Ferster & Skinner, 1957). The ESB-reinforced behavior is different from conventionally reinforced behavior in situations where the rewards follow one another closely in time because the former is controlled by stimuli (the ESB-produced stimuli) that decay rapidly with time, whereas the latter is controlled by stimuli that do not decay rapidly with time. In other situations, with long intervals between rewards or between one reward and the next opportunity to respond, both types of behavior—behavior maintained by ESB reward and behavior maintained by conventional reward—may come under the control of the same kind of stimuli; in these situations, the two types of behavior may be quite similar, even though the rewards remain basically different, in terms of stimulus aftereffects.

### *ESB-Produced Stimuli as Controlling Stimuli*

A given stimulus can serve, simultaneously or at different times, several different functions: it can serve as a secondary reinforcer, as a response instigator, as a cue for responding, and as a discriminative stimulus. By "controlling stimulus" is meant a stimulus that serves both as a response instigator and as a cue; it is assumed in this paper that a conditioned stimulus serves these same two functions.

Rewarding ESB has been found to be an effective conditioned stimulus and also an effective discriminative stimulus in the classical conditioning situation (Hupka, 1970) and an effective conditioned stimulus for an avoidance response in the shuttle box (Mogenson, 1964; Mogenson & Morrison, 1962; Stutz & Asdourian, 1965), which is an instrumental conditioning situation. Furthermore, rewarding ESB has been found to be an effective discriminative stimulus in the operant conditioning (lever-pressing) situation (Butcher & Stutz, 1969; Stutz, Butcher, & Rossi, 1969). Stutz et al. (1969) reported that rewarding ESB was an effective discriminative stimulus when paired with a neutral ESB as an S−, but not when paired with another rewarding ESB, to a different brain area. However, Butcher and Stutz (1969) found that, with the ESB used for reinforcement as well as for the discriminative stimulus, one rewarding ESB (to the septum) can be discriminated from another rewarding ESB (to the hypothalamus).

Valenstein (1964) summarized some studies showing that ESB below reinforcing threshold has cue properties and concluded that an animal "can be trained to respond to the brain stimulation as to any other discriminable stimulus [p. 424]."

Finally, it should be noted that ESBs to nonrewarding parts of the brain have for a long time been used as conditioned and discriminative stimuli in classical and operant conditioning situations (Doty & Giurgea, 1961; Elster, Vandercar, & Schneiderman, 1970; Loucks, 1938; Nielson, Knight, & Porter, 1962). There is no reason to suspect that ESBs to rewarding areas of the brain should be different in this respect.

### *Nature of the ESB Aftereffects*

How are these conditioned or controlling stimuli produced or generated by the ESB? One possibility is based on a special hypothesis as to the nature of the aftereffects of ESB. This hypothesis assumes that the ESB produces a biphasic stimulus event, an initial inhibitory phase followed by a rebound excitatory phase. The initial inhibitory phase constitutes the rewarding aspect of the ESB. The rebound excitatory phase is the source of the stimuli that potentially control the response. When a response leading to a reward occurs while these rebound stimuli are operative, the response will become connected to them.

This hypothesis is similar to the hypothesis of separate pathways advanced by the advocates of the drive-decay theory. But there are two differences. First of all, the drive-decay hypothesis assumes that the second system is a drive system, whereas the associative hypothesis assumes that it is a system for, among other things, generating conditionable stimuli. Second, the drive-decay hypothesis posits that the two systems are anatomically separate, whereas the associative hypothesis posits that they are separate on the temporal dimension.

The recent studies by Ball (1970) and Terman, Terman, and Kling (1970) lend support to the hypothesis of temporally separate systems. Terman et al. demonstrated that the positive effect of ESB intensity on the response rate in the CRF situation is due to the suppression of long interresponse times (IRTs) and an increase in the relative frequency of short IRTs. This could be explained by assuming that increasing the ESB intensity has the effect of increasing the weight of the ESB-produced controlling stimuli, so that the animal responds more quickly to these stimuli as they come up during the rebound phase. With very high intensities, the response rate in some animals decreased, as a result of the suppression of very short IRTs. This could be explained by assuming that the high-intensity ESB increases the duration of the initial reward phase and the delay between the delivery of the ESB and the onset of the rebound phase, which contains the stimuli that control the response.

Ball (1970) presented evidence that the reward component of hypothalamic ESB is temporally separate from the component that gives rise to feeding behavior. The stimuli that produce the feeding response presumably do so because they have innate connections with the response. But the associative hypothesis assumes that these stimuli, or other stimuli in the rebound phase, can also become connected to instrumental responses, such as lever pressing in an operant conditioning box.

### *Six Behavioral Differences*

*Lability.* ESB-maintained behavior is labile in those situations where it is controlled by the stimuli produced by the ESB. It is labile because these stimuli decay rapidly with time.

The rapid rate of extinction following CRF training and the extinction without responding phenomenon (Howarth & Deutsch, 1962) are due to the rapid decay of the ESB-produced controlling stimuli following withdrawal of the ESB reward. In a sense, this is not true extinction: The controlling stimuli, the stimuli connected to the response, are not present, except briefly, during the extinction period; therefore, the stimulus-response connection cannot be broken. The animal stops responding not because the stimulus-response connection is broken but because the controlling stimuli are no longer present or effective.

Herberg (1963) argued that the rapid extinction is an artifact resulting from overtraining. Herberg and Watkins (1966) obtained evidence that the rapid extinction is not due to overtraining as such, but rather to conditioned inhibition generated by repeated extinction and reacquisition sessions; the animal is "trained to extinguish." Both of these studies can be explained by the associative model. In the first study, the long series of 3-minute bouts of self-stimulation may have made time a stimulus for responding, in addition to the ESB-produced stimuli. The prolonged extinction responding may have been due to this time stimulus. This analysis implies that rapid extinction is not an artifact but rather the normal result in situations where the ESB-produced stimuli are not supplemented by other stimuli that promote responding during the extinction period. In the second study, the "training to extinguish" establishes conditioned stimuli for not responding, with the result that the extinction scores become even smaller.

The study by Gibson et al. (1965) showed that increasing the delay of reward, making the lever-press response only one in a chain of responses leading to the ESB reward, leads to an increase in resistance to extinction, to a level comparable to that displayed by food-rewarded animals trained with the same spatio-temporal relation between response and reward. The associative model explains this by assuming that, due to the imposition of a reward delay, environmental stimuli (e.g., the sight of the lever or the noise of the apparatus at which the reward is obtained) or internal stimuli (e.g., the kinesthetic stimuli associated with moving back toward the lever) become connected to the response. These extra controlling stimuli do not, as the ESB-produced stimuli do, decay rapidly during extinction; therefore, extinction responding is augmented.

The results of the Gandelman, Panksepp, and Trowill (1968) experiment, showing that training with lever retraction increases resistance to extinction, are easily explicable in terms of the associative model. The lever, or some stimulus property of the lever, becomes, for the animals in the retraction groups, a controlling stimulus, which, during extinction and in the absence of ESB-produced stimuli, will continue to produce responses and thus increase the resistance to extinction.

The difficulty of shifting from a continuous ESB reinforcement to a partial ESB reinforcement schedule is due to the rapid decay of the ESB-produced stimuli during periods of nonreinforcement. In the typical ESB situation, when a period of nonreinforcement is introduced following a period of continuous reinforcement, the ESB-produced stimuli are the only stimuli that can produce continued responding; if the interval of nonreinforcement is long enough, these stimuli will decay and the responding will cease. In the situation where conventional rewards are used, the controlling stimuli (including perhaps hunger stimuli, kinesthetic stimuli, and environmental stimuli) usually do not decay with time during the nonreinforcement period, so the animal continues responding and obtains further reinforcements. The conventionally reinforced animal, because it continues to respond during the periods of nonreinforcement, has a chance to learn the new reinforcement contingencies, and to learn to respond to new stimuli, such as time (fixed interval [FI] schedule) or number of responses (fixed ratio [FR] schedule); but the ESB-reinforced animal, because it is more likely to stop responding during the long nonreinforcement period, has less of a chance to learn the new reinforcement contingencies or become conditioned to the new stimuli.

This explains why it is not possible to shift directly from continuous ESB reinforcement to a high ratio (e.g., FR 50) or long interval (e.g., FI 8 minutes) schedule, whereas such a shift is possible with conventional rewards (Ferster & Skinner, 1957, p. 48, 139). The shift with ESB animals must be very gradual, starting with short nonreinforcement intervals, so that the responding, controlled by the ESB-produced stimuli, can be carried through to the next reinforcement period. When the new stimuli, arising out of the partial reinforcement schedule, such as time (FI schedule) or response number (FR schedule), begin to gain control of the behavior, then the nonreinforcement interval can be lengthened, because these new stimuli will promote responding during the nonreinforcement periods. If the responding in the ESB animal can be brought under the control of the schedule-related stimuli, instead of the ESB-produced stimuli, then it will come more and more to resemble the responding in the conventionally reinforced animal, since the responding in both groups of animals will be under the same kind of stimulus control. There may, however, still be a difference in the overall behavior pattern of the two groups of animals, in those situations where the offset of reinforcement is a stimulus for not responding (Sadowsky, 1969); the offset of ESB reinforcement may constitute a substantially different stimulus event from the offset of conventional reinforcement.

The establishment of extra controlling stimuli (e.g., the kinesthetic stimuli associated with moving back to the permanent lever) might have been one factor behind the improved partial reinforcement performance observed by Pliskoff et al. (1965) in the retractable lever situation. However, these extra controlling stimuli are evidently not essential for obtaining good partial reinforcement performance with ESB (Brown & Trowill, 1970). Nor is the use of multiple-train rewards essential for good performance (Lenzer & Frommer, 1968, 1971). The most important factor is probably the rate of the shift from CRF to partial reinforcement: the more gradual and careful the shift, the higher the likelihood the animal will learn to respond to the schedule-related stimuli, in place of the ESB-produced stimuli.

The increased resistance to extinction following partial reinforcement, as compared to CRF training (Brown & Trowill, 1970; Culbertson et al., 1966; Herberg, 1963; Pliskoff et al., 1965), is simply due to the difference in stimulus control. In the CRF situation, the controlling stimuli are the ESB-produced stimuli; when the ESB is withdrawn, these stimuli decay and responding ceases. In the partial reinforcement situation, the controlling stimuli are time (FI schedule), number of responses (FR schedule), response rate (FR schedule), and other such stimuli (Ferster & Skinner, 1957). These stimuli do not decay with time following the removal of the ESB. For example, with an FI schedule, the time stimulus, which is controlling the responding, comes up again in extinction, and the animal responds as usual. When no reinforcement is forthcoming, the animal, if it does not know that the time interval has ended, will continue to respond, in expectation of being rewarded (see Pliskoff et al., 1965, Figure 3). Even after the animal has stopped responding once, the time stimulus may come up again and produce another burst of responding. The fact that animals responding on an FI schedule show less resistance to extinction than animals responding on a variable interval (VI) schedule (Brown & Trowill, 1970) can also be understood in terms of differing stimulus control.

Lability in the runway situation is also explained in terms of the rapid decay of the ESB-produced controlling stimuli. When an animal is trained to respond for ESB reward in the Skinner box on a CRF schedule and then in the runway with short intertrial intervals (ITIs), the ESB-produced stimuli will be important in controlling the response. Then when the animal is tested after a long ITI, these controlling stimuli will have dissipated and the running time will be slow. In runway situations using conventional rewards, the reward-produced stimuli do not usually, even with short ITIs, become controlling stimuli; or if they do become connected to the response, it is not to the exclusion of connections between other stimuli and the response, environmental stimuli, and perhaps deprivation-related stimuli. The change in stimulus control resulting from an increase in ITI is minimal; that is, these environmental and deprivation-related stimuli are the same no matter what the duration of the ITI. Therefore, changing duration of ITI does not affect runway performance.

If an animal is trained only with long ITIs, it may, even with hypothalamic reward, fail to perform as well as an animal trained with short ITIs (Seward et al., 1960). But this is not necessarily the case. Scott (1967) reported that animals responding for hypothalamic reward with 15-minute ITIs performed as well or nearly as well as animals with massed trials. The associative model assumes that for an animal to respond rapidly after long ITIs, it must learn to respond to environmental stimuli, instead of the ESB-produced stimuli; once it has learned this, its behavior will be minimally affected by ITI duration. Perhaps some aspect of the Seward et al. (1960) procedure (e.g., extensive pretraining on CRF schedule or training only every second day) made it difficult for the animals in the spaced group to learn to respond to the environmental stimuli, while the Scott (1967) procedure encouraged such conditioning. Scott himself argued that lability in the runway is related to the learning process.

While nonlabile runway behavior can be obtained with hypothalamic reward, previous reports indicate that it cannot be obtained with forebrain reward (Olds, 1956; Scott, 1967; Spear, 1962). This may be explained in associative terms by assuming that some aspect of forebrain reward, not having to do with incentive value, makes it more difficult for the animal to learn to respond, or respond appropriately, to the environmental stimuli. One possibility is that the aftereffect of septal stimulation is more long lasting and therefore reinforces competing responses, as well as the running response (cf. Lindholm & Keesey, 1970). Another possibility is that septal ESB has a larger inhibitory effect on sensory input (cf. Ball, 1967) and consequently makes it more difficult for the response to become connected to the environmental stimuli.

The overnight decrement in performance is also the result of the decay of the ESB-produced controlling stimuli. Priming is effective in reducing or eliminating the decrement because it reinstates the controlling stimuli.

The overnight decrement can be reduced or eliminated by training the animals to respond to environmental stimuli, such as runway stimuli (Scott, 1967), T-maze stimuli (Kornblith & Olds, 1968), discriminative stimuli in a two-compartment shuttle box (Stutz, Lewin, & Rocklin, 1965) or discriminative stimuli in the operant conditioning apparatus (Lenzer & Frommer, 1968). These stimuli, unlike the ESB-produced stimuli, do not decay overnight; they are operative at the beginning of each daily session.

Wasden et al. (1965) reported that the overnight decrement was not displayed by all animals, that whether or not it occurs depends on the brain area involved. In those animals that displayed an overnight decrement, it could be reduced by administering sodium amytal, a fear-reducing drug, no matter whether the animals received ESB to areas in the hypothalamus (Wasden & Reid, 1968) or to areas in the septum, the dorsal hippocampus, or the amygdala (Reid et al., 1970). These results suggest that the overnight decrement is the result of an approach-avoidance conflict resulting from the ambivalent nature of the ESB to certain brain areas. However, the associative model can also account for these data. The overnight decrement and the need for priming are the result of the decay of the ESB-produced controlling stimuli and the absence of any environmental stimuli connected to the response. Animals that do not show any overnight decrement in the runway or the T maze or the operant conditioning box have evidently learned to respond to environmental stimuli; animals that do show the decrement have not learned to do this. The ESB to certain brain areas may inhibit or distort sensory input (cf. Ball, 1967) and thus prevent or hinder the formation of connections between the external stimuli and the response. Animals responding for such ESBs would show the overnight decrement. Sodium amytal may reduce the overnight decrement by reducing the effect of the ESB on incoming sensory input, and thus promoting the formation of connections between the external stimuli and the response.

Reid et al. (1970) reported that the drug was effective on the very first trial following injection, before the animal had experienced the ESB in the drugged condition; this would indicate that the effect of the drug on the overnight decrement is not the result of its effect on the learning process. This finding remains to be confirmed; but assuming that it is true, the associative model could generate several other possible explanations. The drug may make the animal more attentive to certain environmental stimuli that have already, before the drug treatment, been conditioned to the response. Or the drug may reduce the frequency of competing responses and thus increase the effectiveness of the environmental controlling stimuli.

*Insatiability*. The associative model accounts for the insatiability of ESB-maintained behavior in the following way. In the CRF situation the ESB-produced stimuli are the controlling stimuli. Each ESB reproduces these stimuli, thus keeping them at a constant high intensity or amplitude or "weight" (Estes, 1958). The animal is most attentive to these stimuli and thus spends most of its time performing the conditioned response.

Olds (1958) reported that animals responding for septal reward satiate more readily than animals responding for hypothalamic reward. One way to explain this is based on the assumption that the rate of responding for ESB reward is determined mainly by the seizure threshold; the higher the threshold, the less likely a seizure and the higher the rate (Herberg & Blundell, 1969; Newman, 1967; Reid, Gibson, Gledhill, & Porter, 1964; Reid & Porter, 1965). It may be the case that over a prolonged session the septal ESB has the effect of lowering the seizure threshold, which would cause the animal to respond more slowly, in order to avoid a seizure. The hypothalamic ESB may have less of an effect on the seizure threshold and therefore less of an effect on the response rate.

Another possible explanation is based on the hypothesis, already outlined, that the ESB produces a biphasic stimulus event, an initial inhibitory phase constituting the reward aspect of the ESB and a rebound excitatory phase constituting the source of the stimuli that control the response in the CRF situation. The septal ESB, over a prolonged period, may have the effect of increasing the delay between the delivery of the ESB and the onset of the rebound phase; this would lead to a slower rate of responding, since the responding is controlled by the stimuli in the rebound phase. The hypothalamic ESB may be different from the septal ESB in that it has less of an effect, over a prolonged period, on the parameters of the ESB-produced biphasic aftereffect.

*Rate of acquisition*. The rapid rate of acquisition of the lever-press response in the CRF situation is the result of two factors. The first is the absence of a delay between the response and the delivery of the reward. This has mainly to do with the mechanics of the situation; the animal can stay right at the lever, it does not have to move away to collect the reward. The second factor is that the ESB-produced stimuli, because they are reproduced by each ESB, exercise, if the ESB intensity is high enough, almost complete control over the animal's behavior; they keep the animal at the lever so that it does not have the opportunity to sample other stimuli, which might give rise to competing responses, which would retard acquisition.

Fast acquisition can also be obtained with a conventional reward, if the reward is delivered immediately and if the reward produces stimuli that can be conditioned to the response (Panksepp & Trowill, 1967b). This can be taken as evidence that ESB reward is not basically different from conventional rewards (Trowill et al., 1969). On the other hand, it could be argued that this equivalence obtains only between ESB reward and certain conventional rewards (e.g., chocolate milk), not all conventional rewards. For one thing, some conventional rewards (e.g., those pertaining to sex) cannot be delivered immediately, cannot be injected into an animal immediately after an instrumental response. Second, most conventional rewards probably do not produce stimuli suitable for serving as controlling stimuli; they are probably too diffuse and long lasting. Third, the ESB reward, because it does not produce such long-lasting aftereffects, may have less of a reinforcing effect on competing responses than most conventional rewards.

This last point is substantiated by the experiments reported by Lindholm and Keesey (1970). Even though the training and deprivation conditions were identical, even though the stimulus control was the same, animals given ESB reward learned a two-choice brightness discrimination task faster than animals given food reward. The explanation here must be, it seems, in terms of the differing stimulus aftereffects of the two different rewards.

*Dominance*. The associative model accounts for the self-starvation phenomenon (Routtenberg & Lindy, 1965; Spies, 1965) in the following way. In the two-bar preference apparatus (Routtenberg & Lindy, 1965) or in the T maze with massed trials (Spies, 1965) the response for ESB reward is under the control of the ESB-produced stimuli. If the ESB intensity level is high, these controlling stimuli will have a large weight or amplitude. Each ESB the animal receives will reproduce these weighty controlling stimuli. The animal will continue to respond for the ESB, as opposed to food or water, because the stimuli controlling the response for ESB outweigh the stimuli controlling the response for food or water.

As the animal becomes more and more hungry or thirsty, the weight of the hunger or the thirst stimuli will increase. But at the same time the hunger or the thirst will make the corresponding brain system more sensitive, so that the ESB, if it is delivered to some area in the drive-sensitive system, will produce more intense or higher amplitude stimuli. In other words, the ESB-produced stimuli become larger, at the same time that the drive stimuli become larger, and as a result of the same process, namely, deprivation. Therefore, the ESB-produced stimuli may continue to control the behavior, even to the point where the animal starves or dehydrates itself. Or, if the drive stimuli become intense enough, relative to the ESB-produced stimuli, the animal may show a "breakthrough" (Falk, 1961) to responding for food or water.

Increasing the ESB intensity may lead to an increase in the dominance of the response for ESB reward (Mogenson, 1969; Morgan & Mogenson, 1966). This could be explained by assuming that the higher the intensity of the ESB, the greater the weight of the ESB-produced controlling stimuli. However, it could also be explained in terms of an increase in the incentive value of the ESB. There is no doubt that some behavioral changes are due to changes in incentive value. For example, in the T maze with spaced trials, with the choice between water and ESB reward, if the water is made unpalatable by the addition of quinine, the animal will choose the ESB more often than before; this would be explained in terms of a change in incentive value, not a change in the weight of controlling stimuli. Such an explanation would not be inconsistent with the associative model, which nowhere implies that behavior is not affected by changes in reward value. However, the associative model does assume that many behavioral changes that have in the past been explained in terms of changes in reward value are actually the result of changes in the weights of the controlling stimuli.

Moreover, it does not seem that the self-starvation phenomenon can be explained solely in terms of the idea of reward value; this is where the associative analysis comes in. On the basis of this analysis, one would predict that in a T maze with spaced trials, with the behavior firmly under the control of environmental and thirst or hunger stimuli, there would be no self-starvation or self-dehydration. This is in fact what the Deutsch et al. (1964) experiment suggests.

*Secondary reinforcement.* Mogenson (1965) reported that secondary reinforcement effects could not be demonstrated with nondeprived animals trained with ESB as the primary reinforcement. This is so, according to the associative model, because the stimuli controlling the response for the secondary reinforcement, namely, the ESB-produced stimuli, are not present during the test for secondary reinforcement, and so the animal does not respond. The neutral stimulus may possess secondary reinforcement properties, but these may remain latent, because of the absence of the controlling stimuli. If the controlling stimuli are reintroduced during the test for secondary reinforcement, by giving the animal random, noncontingent ESBs, then the secondary reinforcement properties can be made manifest (Fantl & Schuckman, 1967).

DiCara (see Footnote 3) reported that secondary reinforcement can be demonstrated in the food-deprived animal, without any ESBs during the test period. One possible explanation, in associative terms, is that the food deprivation produces, in those animals with electrodes in the hunger system, stimuli that resemble the ESB-produced controlling stimuli. These hunger stimuli can, then, support the response during the test for secondary reinforcement.

## Some Comparisons between the Theories

### *Associative versus Drive Decay*

There are many studies (see the first section of this paper; see also Trowill et al., 1969) that show that in certain situations, with certain training and deprivation conditions, ESB-maintained behavior is not labile. One of the most important of these studies is that by Pliskoff et al. (1965), which showed that nondeprived animals can be trained to respond for posterior hypothalamic reward on difficult schedules of reinforcement. Gallistel (1969a, p. 720) offered two possible explanations of this finding. First of all, he argued that the multiple-train rewards used by Pliskoff et al. could have produced prolonged motivational aftereffects, which would act to sustain responding during the nonreinforcement periods. However, the experiments by Lenzer and Frommer (1968, 1971), showing that good performance on an operant discrimination task can be established with only a single ESB train for each S+ period, do not support this explanation. Second, Gallistel (1969a) suggested that once an animal has learned to respond on a schedule with long intervals of nonreinforcement, it "may no longer require a high-level aftereffect in order to continue responding, and may make response bursts that outlast the aftereffect of BSR [brain-stimulation reward] [p. 720]." However, it is not clear what this means. If an animal needs to be motivated in order to respond, how can it respond without being motivated? To be sure, it learns to respond, but in the absence of motivation? The associative model assumes that the controlling stimuli serve as the response instigators, that is, as the source of motivation. Animals can learn to respond well on difficult reinforcement schedules because they can learn to respond to the stimuli connected with the schedule. The learning involves the transferring of response control from the ESB-produced stimuli to the stimuli associated with the schedule. Gallistel would have to assume that in the early stages of learning the motivation is in the form of drive, while in the later stages it is in the form of incentive motivation; this would be a relatively uneconomical explanation.

Sonderegger and Rose (1970) and Milner (1970) have also argued that the drive-decay theory can be saved by assuming, in Milner's words, that "some motivational systems can be conditioned to environmental stimuli [p. 398]." But by "motivational systems" Milner cannot mean drive systems; as Bolles (1967) has said, "there is little indication that satiated animals can be made hungry or thirsty by the presentation of external stimuli [p. 309]." Milner must mean the systems having to do with incentive motivation. But here again one would have to assume that the motivation for ESB-maintained behavior is sometimes in the form of drive and sometimes in the form of incentive motivation. The associative model assumes that the controlling stimuli, whether they be the ESB-produced stimuli or the environmental stimuli, serve to instigate and guide the response, and the instigation is always of the same form.

### *Associative versus Conflict*

The conflict hypothesis assumes that the longer the time since the last ESB, the greater the relative strength of the aversive component, the greater the approach-avoidance conflict, and the greater the likelihood that the animal will not respond for ESB again. When an animal receiving an ambivalent ESB pauses during a prolonged session, the conflict should be increasing during the pause. If the pause is long enough, the animal should avoid the ESB totally and never resume responding. But the data presented by Olds (1958) show that animals receiving septal ESB, which is known to be ambivalent (Olds & Olds, 1963), continually return to the lever and resume responding, even after very long pauses; there does not seem to be any active avoidance of the ESB in this situation.

Lenzer and Frommer (1968, 1971) have shown that animals receiving ESB to ambivalent areas of the brain, for example, the septum and the far lateral hypothalamus, respond without lability in the operant discrimination situation. These experiments suggest that in those situations where the external controlling stimuli are clear-cut and easily conditioned, animals receiving ESB to ambivalent areas will perform very much like animals receiving

ESB to more purely rewarding areas. The experiment by Scott (1967) lends support to this point. Some animals received ambivalent ESB, as indicated by histological examination and by tests for escape behavior, and yet ran very well in the runway with spaced trials and did not show any overnight decrement. Again this suggests that lability is a function of the degree of conditioning to stable environmental stimuli, and not a function of the degree of ambivalence associated with the ESB. The fact that the ESB in some animals has an aversive component may be coincidental or secondary to the effect on learning.

### *Associative versus Incentive*

There are several experiments that the incentive model, as formulated by Trowill et al. (1969), apparently cannot account for. Gallistel (1966) trained animals to run in two runways in strict alternation, with a 6-second ITI. When the ESB in one of the two goal boxes was set at a lower current intensity level than the ESB in the other goal box, the animal ran faster in the runway with the lower ESB. This was despite the fact that in the T-maze situation they chose the arm with the higher ESB, indicating that the higher intensity ESB was more rewarding, or they showed no preference. If the "motivational properties [of ESB] are inherent in its reinforcement properties in the form of incentive motivation [Trowill et al., 1969, p. 264]," then the animal should run faster in the runway with the high-intensity ESB. The runway with the higher level of reinforcement should also have the higher level of motivation.

Gallistel (1966) accounted for this experiment in terms of the drive-decay theory and concluded that "the stimulation in electrical self-stimulation has a motivating effect which is independent of its rewarding effect [p. 101]." However, the associative model could also account for the data. The animal running in the runway with the lower ESB has just received the higher ESB, as a reward for running the other runway. Therefore, the ESB-produced stimuli, which, with a 6-second ITI, are controlling the response in both runways, are more intense or weighty in the runway with the lower ESB. Therefore, the running is faster in that runway.

Gandelman et al. (1968) reported that responding during extinction was augmented following training with lever retraction. Giving the animals free ESBs during the period of lever retraction did not significantly reduce the effect of this training procedure. This result goes against the hypothesis that increased resistance to extinction after training with lever retraction and after training on partial reinforcement schedules is due to a learning to respond to the "cues of frustration" or nonreward (Trowill et al., 1969). It can, however, be explained by assuming that the training with lever retraction establishes the lever as a controlling stimulus, which can, during extinction and in the absence of the ESB-produced controlling stimuli, sustain or promote responding.

The experiment by Johnson, Lobdell, and Levy (1969)—which shows that the frustration generated by withholding the ESB reinforcement in the two-box situation decays very rapidly, even, apparently, more rapidly than the frustration generated by withholding conventional reinforcement (Amsel & Roussel, 1952; MacKinnon & Amsel, 1964)—throws further doubt on the frustration interpretation of the increased resistance to extinction following partial ESB reinforcement. Trowill et al. (1969) assume that the partial reinforcement effect is due to the fact that the animals have learned to respond to the cues associated with frustration. But if these stimuli decay rapidly with time, they will not be effective in sustaining responding for a very long time. Graphs of extinction responding following partial ESB reinforcement (e.g., those in the report by Pliskoff et al., 1965) show that the animals continue to respond, at first continuously and then in bursts, for many minutes after the beginning of the extinction period. The Johnson et al. (1969) study suggests that this responding is not controlled by the cues of frustration. The associative model assumes that the responding is controlled by the stimuli associated with the particular schedule the animal is trained on —for example, the time stimulus on an FI schedule or the number of responses and/or the rate of responding stimulus on the FR schedule—and also perhaps by certain environmental and kinesthetic stimuli.

Although there are certain differences between them, the associative model and the incentive model nevertheless have many features in common, and the associative model can even be viewed from the standpoint of an incentive theory of motivation. A controlling stimulus, in its function as a response instigator, is an incentive-motivational stimulus, according to Bindra's (1968) definition. The associative theory of rewarding brain stimulation, although it emphasizes stimulus-response relations, could be incorporated into a broader theory, such as the stimulus-response, cybernetic theory developed by Estes (1969, 1970), which treats motivational phenomena essentially from an incentive-motivational point of view.

If it is assumed that the ESB-produced stimuli can become controlling stimuli, then the experiments by Gallistel (1966, 1969b) can be incorporated into an incentive framework. The ESB-produced stimuli derive their response-instigating properties from their association with the ESB as reinforcement; by definition, they are incentive-motivational stimuli. But once the ESB-produced stimuli have become conditioned, they can be manipulated independently of the ESB reinforcement. But the fact that the two aspects can be manipulated separately does not imply that they are separate in the final analysis: the motivational aspect is ultimately dependent on the reinforcement aspect.

The associative analysis suggests that ESB reward is, in certain fundamental respects, different from conventional rewards. The uniqueness of ESB can be illustrated by considering the relation between drive manipulations and instrumental behavior. Bindra (1968) argued that drive does not affect instrumental responding directly, but indirectly, through its effect on incentive motivation. His argument, based on the study by Mendelson (1966), is that drive facilitates instrumental behavior by facilitating consummatory behavior and the likelihood that the animal will collect the reward and be reinforced. In the ESB situation, however, drive may directly facilitate instrumental responding. The drive, by sensitizing the neural tissue underlying the electrode, may augment the amplitude or the weight of the ESB-produced controlling stimuli. Since these stimuli serve (in some situations) both to instigate and to guide the behavior, the drive, by increasing their weight, will have increased both the level of incentive motivation and the degree of response guidance. In this way drive could have a direct effect on instrumental behavior, direct in the sense that it is not mediated by an effect on consummatory behavior.

## REFERENCES

Amsel, A., & Roussel, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *Journal of Experimental Psychology,* 1952, **43**, 363–368.

Ball, G. G. Electrical self-stimulation of the brain and sensory inhibition. *Psychonomic Science,* 1967, **8**, 489–490.

Ball, G. G. Hypothalamic self-stimulation and feeding: Different time functions. *Physiology and Behavior,* 1970, **5**, 1343–1346.

Ball, G. G., & Adams, D. W. Intracranial stimulation as an avoidance or escape response. *Psychonomic Science,* 1965, **3**, 39–40.

Bindra, D. Neuropsychological interpretation of the effects of drive and incentive-motivation on general activity and instrumental behavior. *Psychological Review,* 1968, **75**, 1–22.

Black, R. W. Shifts in magnitude of reward and contrast effects in instrumental and selective learning: A reinterpretation. *Psychological Review,* 1968, **75**, 114–126.

Black, R. W. Incentive motivation and the parameters of reward in instrumental conditioning. *Nebraska Symposium on Motivation,* 1969, **17**, 85–137.

Bolles, R. C. *Theory of motivation.* New York: Harper & Row, 1967.

Brady, J. V., & Conrad, D. Some effects of limbic system self-stimulation upon conditioned emotional behavior. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 128–137.

Brodie, D. A., Moreno, O. M., Malis, J. L., & Boren, J. J. Rewarding properties of intracranial stimulation. *Science,* 1960, **131**, 929–930.

Brown, S., & Trowill, J. A. Lever-pressing performance for brain stimulation on F-I and V-I schedules in a single-lever situation. *Psychological Reports,* 1970, **26**, 699–706.

Butcher, R. E., & Stutz, R. M. Discriminability of rewarding subcortical brain shock. *Physiology and Behavior,* 1969, **4**, 885–887.

Cotton, J. W., & Lewis, D. J. The effect of intertrial interval on the acquisition and extinction of a running response. *Journal of Experimental Psychology,* 1957, **54**, 15–20.

Culbertson, J. L., Kling, J. W., & Berkley, M. A. Extinction responding following ICS and food reinforcement. *Psychonomic Science,* 1966, **5**, 127–128.

Deutsch, J. A. Learning and electrical self-stimulation of the brain. *Journal of Theoretical Biology,* 1963, **4**, 193–214.

Deutsch, J. A., Adams, D. W., & Metzner, R. J. Choice of intracranial stimulation as a function of delay between stimulations and strength of competing drive. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 241–243.

Deutsch, J. A., & Deutsch, D. *Physiological psychology.* Homewood, Ill.: Dorsey, 1966.

Deutsch, J. A., & DiCara, L. Hunger and extinction in intracranial self-stimulation. *Journal of Comparative and Physiological Psychology,* 1967, **63**, 344–347.

Deutsch, J. A., & Howarth, C. I. Some tests of a theory of intracranial self-stimulation. *Psychological Review,* 1963, **70**, 444–460.

Doty, R. W., & Giurgea, C. Conditioned reflexes established by coupling electrical excitation of two cortical areas. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning.* Oxford: Blackwell, 1961.

Dunham, P. J. Contrasted conditions of reinforcement: A selective critique. *Psychological Bulletin,* 1968, **69**, 295–315.

Elder, S. T., Montgomery, N. P., & Rye, M. M. Effects of food deprivation and methamphetamine on fixed-ratio schedules of intracranial self-stimulation. *Psychological Reports,* 1965, **16**, 1225–1233.

Elster, A. J., Vandercar, D. H., & Schneiderman, N. Classical conditioning of heart rate discriminations using subcortical electrical stimulation as conditioned and unconditioned stimuli. *Physiology and Behavior,* 1970, **5**, 503–508.

Estes, W. K. A study of motivating conditions necessary for secondary reinforcement. *Journal of Experimental Psychology,* 1949, **39**, 306–310.

Estes, W. K. Stimulus-response theory of drive. *Nebraska Symposium on Motivation,* 1958, **6**, 35–69.

Estes, W. K. New perspectives on some old issues in association theory. In N. J. MacKintosh & W. K. Honig (Eds), *Fundamental issues in associative learning.* Halifax: Dalhousie University Press, 1969.

Estes, W. K. *Learning theory and mental development.* New York: Academic Press, 1970.

Falk, J. L. Septal stimulation as a reinforcer of and an alternative to consumatory behavior. *Journal of the Experimental Analysis of Behavior,* 1961, **4**, 213–217.

Fantl, L., & Schuckman, H. Lateral hypothalamus and hunger: Response to a secondary reinforcer with and without electrical stimulation. *Physiology and Behavior,* 1967, **2**, 355–357.

Ferster, C. B., & Skinner, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

Gallistel, C. R. Electrical self-stimulation and its theoretical implications. *Psychological Bulletin,* 1964, **61**, 23–34.

Gallistel, C. R. Motivating effects in self-stimulation. *Journal of Comparative and Physiological Psychology,* 1966, **62**, 95–101.

Gallistel, C. R. Intracranial stimulation and natural reward: Differential effects of trial spacing. *Psychonomic Science,* 1967, **9**, 167–168.

Gallistel, C. R. The incentive of brain-stimulation reward. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 713–721. (a)

Gallistel, C. R. Self-stimulation: Failure of pretrial stimulation to affect rat's electrode preference. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 722–729. (b)

Gandelman, R., Panksepp, J., & Trowill, J. The effect of lever retraction on resistance to extinction of a response rewarded with electrical stimulation of the brain. *Psychonomic Science,* 1968, **10**, 5–6.

Gibson, W. E., Reid, L. D., Sakai, M., & Porter, P. B. Intracranial reinforcement compared with sugar-water reinforcement. *Science,* 1965, **148**, 1357–1359.

Herberg, L. J. Determinants of extinction in electrical self-stimulation. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 686–690.

Herberg, L. J., & Blundell, J. E. Epileptiform seizures and the rat's choice of stimulus duration in electrical self-stimulation. *Journal of Comparative and Physiological Psychology,* 1969, **67**, 442–447.

Herberg, L. J., & Watkins, P. J. The effect of overtraining and repeated extinctions on speed of extinction in electrical self-stimulation. *Quarterly Journal of Experimental Psychology,* 1966, **56**, 75–78.

Howarth, C. I., & Deutsch, J. A. Drive decay: The cause of fast "extinction" of habits learned for brain stimulation. *Science,* 1962, **137**, 35–36.

Hupka, R. B. Electrical stimulation of the septum and hypothalamus as conditioned stimuli in the rabbit. *Physiology and Behavior,* 1970, **5**, 1355–1363.

Huston, J. P. Reinforcement reduction: A method for training ratio behavior. *Science,* 1968, **159**, 444.

Johnson, R. N., Lobdell, P., & Levy, R. S. Intracranial self-stimulation and the rapid decline of frustrative nonreward. *Science,* 1969, **164**, 971–972.

Keesey, R. E., & Goldstein, M. D. Use of progressive fixed-ratio procedures in the assessment of intracranial reinforcement. *Journal of the Experimental Analysis of Behavior,* 1968, **11**, 293–301.

Kent, E., & Grossman, S. P. Evidence for a conflict interpretation of anomalous effects of rewarding brain stimulation. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 381–390.

Kling, J. W., & Berkley, M. A. Electrical brain stimulation and food reinforcement in discrimination and generalization situations. *Journal of Comparative and Physiological Psychology,* 1968, **65**, 507–511.

Kling, J. W., & Matsumiya, Y. Relative reinforcement values of food and intracranial stimulation. *Science*, 1962, **135**, 668–670.

Kornblith, C., & Olds, J. T-maze learning with one trial per day using brain stimulation reinforcement. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 488–491.

Lenzer, I. I., & Frommer, G. P. Successive sensory discriminative behavior maintained by intracranial self-stimulation reinforcement. *Physiology and Behavior*, 1968, **3**, 345–349.

Lenzer, I. I., & Frommer, G. P. Successive sensory discriminative behavior maintained by forebrain self-stimulation reinforcement. *Psychonomic Science*, 1971, **23**, 88–90.

Lindholm, E. P., & Keesey, R. E. Faster rates of discrimination learning with centrally elicited reinforcement than with food. *Journal of Comparative and Physiological Psychology*, 1970, **72**, 318–327.

Logan, F. A. *Incentive: How the conditions of reinforcement affect the performance of rats.* New Haven: Yale University Press, 1960.

Loucks, R. B. Studies of neural structures essential for learning. II. The conditioning of salivary and striped muscle responses to faradization of cortical sensory elements and action of sleep upon such mechanisms. *Journal of Comparative Psychology*, 1938, **25**, 315–332.

MacKinnon, J. R., & Amsel, A. Magnitude of the frustration effect as a function of confinement and detention in the frustrating situation. *Journal of Experimental Psychology*, 1964, **67**, 468–474.

McIntire, R. W., & Wright, J. E. Differences in extinction in electrical brain stimulation under traditional procedures of reward presentation. *Psychological Reports*, 1965, **16**, 909–913.

Mendelson, J. The role of hunger in T-maze learning for food by rats. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 341–353.

Miller, N. E. Implications for theories of reinforcement. In D. E. Sheer (Ed.), *Electrical stimulation of the brain*. Austin: University of Texas Press, 1961. (a)

Miller, N. E. Some recent studies of conflict behavior and drugs. *American Psychologist*, 1961, **18**, 12–24. (b)

Milner, P. M. *Physiological psychology*. New York: Holt, Rinehart & Winston, 1970.

Mogenson, G. J. Avoidance responses to rewarding brain stimulation: Replication and extension. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 465–467.

Mogenson, G. J. An attempt to establish secondary reinforcement with rewarding brain stimulation. *Psychological Reports*, 1965, **16**, 163–167.

Mogenson, G. J. Effects of drugs on the preference between electrical stimulation of the lateral hypothalamus and water. *Psychonomic Science*, 1969, **17**, 13–14.

Mogenson, G. J., & Morrison, M. J. Avoidance responses to reward stimulation of the brain. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 691–694.

Morgan, C. W., & Mogenson, G. J. Preference of water-deprived rats for stimulation of the lateral hypothalamus rather than water. *Psychonomic Science*, 1966, **6**, 337–338.

Newman, B. L. Behavioral effects of electrical self-stimulation of the septal area and related structures in the rat. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 340–346.

Newman, B. L. EEG correlates of self-stimulation. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 516–520.

Nielson, H. C., Knight, J. M., & Porter, P. E. Subcortical conditioning, generalization and transfer. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 168–173.

Olds, J. Physiological mechanisms of reward. *Nebraska Symposium on Motivation*, 1955, **3**, 73–139.

Olds, J. Runway and maze behavior controlled by basomedial forebrain stimulation in the rat. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 507–512.

Olds, J. Satiation effects in self-stimulation of the brain. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 675–678.

Olds, J., & Milner, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, 1954, **47**, 419–427.

Olds, M. E., & Olds, J. Approach-avoidance analysis of rat diencephalon. *Journal of Comparative Neurology*, 1963, **120**, 259–295.

Panksepp, J., Gandelman, R., & Trowill, J. A. The effect of intertrial interval on running performance for ESB. *Psychonomic Science*, 1968, **13**, 135–136.

Panksepp, J., & Trowill, J. A. Intraoral self injection: I. Effects of delay of reinforcement on resistance to extinction and implications for self-stimulation. *Psychonomic Science*, 1967, **9**, 405–406. (a)

Panksepp, J., & Trowill, J. A. Intraoral self injection: II. The simulation of self-stimulation phenomena with a conventional reward. *Psychonomic Science*, 1967, **9**, 407–408. (b)

Panksepp, J., & Trowill, J. A. Extinction following intracranial reward: Frustration or drive decay? *Psychonomic Science*, 1968, **12**, 173–174.

Panksepp, J., & Trowill, J. A. Positive and negative contrast effects with electrical stimulation of the brain. *Physiology and Behavior*, 1969, **4**, 173–175.

Panksepp, J., & Trowill, J. A. Positive incentive contrast with rewarding electrical stimulation of the brain. *Journal of Comparative and Physiological Psychology*, 1970, **70**, 358–363.

Pliskoff, S. S., & Hawkins, T. D. Test of Deutsch's drive-decay theory of rewarding self-stimulation of the brain. *Science*, 1963, **141**, 823–824.

Pliskoff, S. S., Wright, J. E., & Hawkins, T. D. Brain stimulation as a reinforcer: Intermittent schedules. *Journal of the Experimental Analysis of Behavior,* 1965, **8**, 75–88.

Porter, J. M. Extinction of an acquired response as a function of the interval between successive nonrewarded trials. *Journal of Comparative Psychology,* 1938, **26**, 261–270.

Premack, D. On some boundary conditions of contrast. In J. T. Tapp (Ed.), *Reinforcement and behavior.* New York: Academic Press, 1969.

Quartermain, D., & Webster, D. Extinction following intracranial reward: The effect of delay between acquisition and extinction. *Science,* 1968, **159**, 1259–1260.

Ray, O. S., Hine, B., & Bivens, L. W. Stability of self-stimulation responding during long test sessions. *Physiology and Behavior,* 1968, **3**, 161–164.

Reid, L. D., Gibson, W. E., Gledhill, S. M., & Porter, P. B. Anticonvulsant drugs and self-stimulating behavior. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 353–356.

Reid, L. D., & Porter, P. B. Reinforcement from direct electrical stimulation of the brain. *Rocky Mountain Psychologist,* 1965, **1**, 3–22.

Reid, L. D., Wasden, R. E., & Courtney, R. J. Reinforcing limbic system stimulation and sodium amytal. *Psychonomic Science,* 1970, **18**, 47–48.

Routtenberg, A., & Lindy, J. Effects of the availability of rewarding septal and hypothalamic stimulation on bar pressing for food under conditions of deprivation. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 158–161.

Sadowsky, S. Discriminative responding on associated mixed and multiple schedules as a function of food and ICS reinforcement. *Journal of the Experimental Analysis of Behavior,* 1969, **12**, 933–945.

Schlosberg, H., & Pratt, C. H. The secondary reward value of inaccessible food for hungry and satiated rats. *Journal of Comparative and Physiological Psychology,* 1956, **49**, 149–152.

Scott, J. W. Brain stimulation reinforcement with distributed practice: Effects of electrode locus, previous experience, and stimulus intensity. *Journal of Comparative and Physiological Psychology,* 1967, **63**, 175–183.

Seward, J. P., Uyeda, A., & Olds, J. Resistance to extinction following cranial self-stimulation. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 294–299.

Seward, J. P., Uyeda, A. A., & Olds, J. Reinforcing effect of brain stimulation on runway performance as a function of interval between trials. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 224–228.

Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *Journal of Experimental Psychology,* 1949, **39**, 511–526.

Sidman, M., Brady, J. V., Boren, J. J., Conrad, D. G., & Schulman, A. Reward schedules and behavior maintained by intracranial self-stimulation. *Science,* 1955, **122**, 830–831.

Sonderegger, T. B., & Rose, G. H. Approach gradients obtained through intracranial stimulation of the medial forebrain bundle in rats. *Journal of Comparative and Physiological Psychology,* 1970, **71**, 52–58.

Spear, N. E. Comparison of the reinforcing effect of brain stimulation on Skinner box, runway, and maze performance. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 679–684.

Spies, G. Food versus intracranial self-stimulation reinforcement in food-deprived rats. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 153–157.

Stutz, R. M., & Asdourian, D. Positively reinforcing brain shock as a CS in the acquisition and extinction of a shuttle box avoidance response. *Psychonomic Science,* 1965, **3**, 191–192.

Stutz, R. M., Butcher, R. E., & Rossi, R. Stimulus properties of reinforcing brain shock. *Science,* 1969, **163**, 1081–1082.

Stutz, R. M., Lewin, I., & Rocklin, K. W. Generality of "drive-decay" as an explanatory concept. *Psychonomic Science,* 1965, **2**, 127–128.

Terman, J. S., Terman, M., & Kling, J. W. Some temporal properties of intracranial self-stimulation. *Physiology and Behavior,* 1970, **5**, 183–191.

Terman, M., & Kling, J. W. Discrimination of brightness differences by rats with food or brain-stimulation reinforcement. *Journal of the Experimental Analysis of Behavior,* 1968, **11**, 29–37.

Trowill, J. A., Panksepp, J., & Gandelman, R. An incentive model of rewarding brain stimulation. *Psychological Review,* 1969, **76**, 264–283.

Valenstein, E. S. Problems of measurement and interpretation with reinforcing brain stimulation. *Psychological Review,* 1964, **71**, 415–437.

Wasden, R. E., & Reid, L. D. Intracranial stimulation: Performance decrements and a fear-reducing drug. *Psychonomic Science,* 1968, **12**, 117–118.

Wasden, R. E., Reid, L. D., & Porter, P. B. Overnight performance decrement with intracranial reinforcement. *Psychological Reports,* 1965, **16**, 653–658.

Wetzel, M. C. Self-stimulation aftereffects and runway performance in the rat. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 673–678.

Wilson, W., Weiss, E. J., & Amsel, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *Journal of Experimental Psychology,* 1955, **50**, 51–60.

(Early publication received October 28, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 2, 119–128

# SELF-MUTILATING BEHAVIOR[1]

DAVID LESTER[2]

*Stockton State College, Pomona, New Jersey*

The varieties of self-multilating behavior are noted, and studies that allow estimates of their prevalence are reviewed, with special focus on head banging, trichotillomania, wrist cutting, and self-castration. The relationship between the study of self-mutilation and the study of suicidal behavior is discussed, and it is noted that research would benefit from a closer interaction between them. Several areas for research are identified: epidemiological studies of the prevalence of self-mutilating behavior, research designed to verify the clinical generalizations about self-mutilators, and studies of the relationship among different kinds of self-mutilating behavior and between self-mutilating behavior and suicidal behavior.

Although self-mutilation is common, its prevalence is often underestimated by psychologists. Self-mutilation among psychiatric patients is well known to medical personnel. Head banging is common among children, hair pulling, lip biting, and skin mutilation are frequently found among the feeble minded, and Betts (1964) and Mintz (1964) have reported examples of self-cannibalism. Mutilation has also been part of common behavior and rituals in many societies (Montagu, 1946). Breast mutilation was stressed by the Greeks in their accounts of Amazon women (Schechter, 1962). Other examples of mutilation as a ritual or behavior in society include deforming of head and foot; piercing and stretching earlobes, nostrils, and lips; scarifying and pigmenting skin; stretching the neck; cutting holes in the head (trephining); chipping, filing, and drilling of teeth; circumcising; and subincising. Self-mutilation has also been associated with religions (Menninger, 1938), for example, the Skoptsi in Russia in the eighteenth century who castrated themselves. Some particular cases of self-mutilation are well known, such as the removal by Van Gogh of the lower part of his outer ear (Lubin, 1961). Self-mutilating behavior is also common in the population at large. Many of us bite nails, pull hair, tug at ear lobes, shave, pierce ears, cut hair and nails, etc. Some instances of self-mutilation are motivated by secondary gain and are labeled *malingering*. This is especially common in soldiers, who sometimes self-inflict wounds to avoid duty (Roper-Hall, 1950).

Self-mutilating behavior is of interest not only as a behavior that merits study but also because of its relationship to suicidal behavior. The study of suicide is being pursued vigorously today, and self-mutilating behavior is relevant, since it is an example of a partially self-destructive behavior that allows the person to live—*focal suicide*, to use Menninger's (1938) term. It forms an important link in the study of self-destructive behavior since it is more directly destructive than smoking or heavy drinking but less directly destructive than suicide attempts.

For Menninger, self-mutilation is a suicidal act in which the suicidal impulse is concentrated upon a part of the self as a substitute for the whole. When describing the suicidal impulse, Menninger noted three components: the desire to hurt others, the desire to be punished, and the desire to escape from an unbearable situation. In self-mutilation, the first two components are apparent. The patient usually is angry at some person and desires to hurt that person. He also feels the need to be punished for some actual or imagined crime or sin. The act of self-mutilation makes the significant other distressed and causes pain to the self; thus it satisfies two motives. Menninger noted that the neurotic rarely mutilates himself irrevocably. He sometimes resorts to symbolic and substitute forms of self-mutilation (such as impotence) and sometimes mutilates himself with the aid of a second party (e.g., through unnecessary surgery). This contrasts the functioning of the neurotic's ego to that of the psychotic, who is more likely to engage in more extreme and bizarre self-mutilations.

---

[1] Work on this paper was initiated while the author was at the Suicide Prevention and Crisis Service in Buffalo, New York.

[2] Requests for reprints should be sent to David Lester, Division of Social Sciences, Stockton State College, Pomona, New Jersey 08240.

The aim of this review is to survey and summarize the literature on self-mutilation. It suggests the direction for future research, and it points out the lacunae in present knowledge. Finally, the implications for the study of suicide are examined.

### *Estimates of Prevalence*

Several studies have reported the incidence of self-mutilating behavior in small, specialized populations. Phillips and Muzaffer (1961) found that 4.3% of psychiatric patients mutilated themselves (mainly by wrist cutting), the incidence in females being three times that in males. Ressman and Butterworth (1952) found that 2.1% of the population of a school for the feeble minded mutilated their skin (mainly by biting and chewing their arms and hands). Clark and Campbell (1948) noted that about 2% of gunshot wounds among soldiers were probably self-inflicted. Dermatological clinics in Munich and at the University of Bonn have reported the proportion of self-mutilators in the population at large to be .06% and .34%, respectively (Olsen, 1967).

Wechsler (1931) noted that nail biting was not apparent in children under 3 years of age. From 3 to 12 years the incidence rose to about 45% and thereafter declined. Levy and Patrick (1928) found the incidence of head banging to be 2.6% and 3.8% in the babies seen in two small clinics in Illinois. Kravitz, Rosenthal, Teplitz, Murphy, and Lesser (1960) found head banging in 3.6% of children seen in private practice. De Lissovoy (1961) followed a sample of babies born in one hospital for 20–30 months and found head banging in 22% of the male children and in 7% of the female children.

Systematic epidemiological studies of the incidence of self-mutilating behavior are needed. It is necessary to estimate the incidence of the behavior both in the general population and in specific groups (such as the retarded, alcoholics, etc.). In the studies reviewed above, the demographic characteristics of the samples were not given, and so the reliability of the percentages is difficult to estimate. It is probably safe to say that the estimates above represent the right order of magnitude of the phenomenon in the groups examined.

### *Research on Prisoners*

Claghorn and Beto (1967) compared schizophrenic male prisoners who were self-mutilators with those who were not. They hypothesized that self-mutilation was a response to frustration, but they failed to find differences in the incidence of crimes of violence between the groups and felt that this fact argued against their hypothesis. They found a high incidence of self-mutilation among Latin American prisoners. They argued that this higher incidence reflected the greater frustration of this group when imprisoned, since, supposedly, the Latin American is traditionally oriented to the individuality and uniqueness of man and would thus be expected to adapt less well to regimentation in prison.

McKerracher, Loughnane, and Watson (1968) studied feeble-minded female prisoners diagnosed as psychopaths. The sample was small and biased due to the difficulty of finding patients who could complete the required tests. McKerracher et al. felt that the self-mutilators were cutting themselves in response to boredom (hence the increase in acts of self-mutilation on weekends, when the patients had less regulated schedules) and in response to frustration. Since self-mutilators did not rate themselves as more bored, tense, or anxious than the nonmutilators, we may agree with McKerracher et al. that the difference between them lies in their *responses* to boredom. The self-mutilators were more phobic, obsessional, and preoccupied with bodily complaints than the nonmutilators, but did not differ in IQ.

Panton (1962) compared the MMPI scores of male self-mutilating prisoners with nonmutilating prisoners. The self-mutilators appeared to be more hostile, to have more inner turmoil, and to be more maladjusted. Panton also claimed that the self-mutilators were suffering from personality disorders rather than psychoses, in spite of the fact that the self-mutilators had significantly higher scores on the *Sc* scale and on an MMPI scale of psychotic tendency. Panton felt that the nonmutilators tended to have a hysterical reaction to stress rather than a self-destructive reaction. He also felt that the self-mutilators were not suicidal in view of the fact that they did not have elevated *D* scores on the MMPI. In a recent review of relevant studies, however, Lester (1970c) found no evidence that the *D* score was related to suicidal tendencies.

### *Research on Schizophrenic Children*

Green (1967) hypothesized that self-mutilation in schizophrenic children was associated with a history of head banging, and he found this to be so in both boys and girls.[3] In a subsequent study on these same children, Green (1968) showed that self-mutilation was significantly associated with experience of physical abuse in the first 2 years of life, yet experience of physical abuse was not associated with the occurrence of infantile head banging. Green therefore considered experience of physical abuse and head banging to be independent precursors of self-mutilation in schizophrenic children.

Green (1968) argued that the schizophrenic child learns self-mutilation. The child is neglected by the parent and suffers from stimulus deprivation. Rocking becomes the first adaptive response in order to increase sensory input and to provide better ego-boundary delineation. The child accidentally bangs his head while rocking, and the head banging becomes associated with the pleasure of rocking. The mother frequently abuses the child, especially as a result of the head banging (which can be noisy, a nuisance, and a worry to the mother), and the child thereby obtains attention and contact during this abuse. The child learns to hurt himself in order to reexperience the pleasurable elements of the parental attack in the absence of normal parental contact. The painful elements of self-mutilating behavior become subordinated to the pleasurable components.

Shodell and Reiter (1968) argued that the ability of schizophrenic children to verbalize enables them to use language as a substitute for action when releasing their tension and frustration and to better understand the environment (and so become less frustrated). They found that, indeed, verbal schizophrenic children were significantly less likely to mutilate themselves than nonverbal schizophrenic children. However, Shodell and Reiter failed to show that age was not a confounding variable in this association.

## SPECIFIC KINDS OF SELF-MUTILATING BEHAVIOR[4]

### *Head Banging*

Two kinds of head banging have been identified: In autoerotic head banging, the child bangs his head until he exhausts himself and falls asleep. Tantrum head banging, which begins when the child is able to walk, is a reaction to being thwarted and frustrated. The child throws himself to the ground, thrashes, and bangs his head. This behavior (occasionally found in adults) is not rhythmical or absorbing and leaves the child exhausted, weakened, and fearful. Most studies of head banging, however, have focused on autoerotic head banging.

De Lissovoy (1961) noted that head bangers were significantly more likely to be last borns than firstborns. In 56% of the children, the mothers noted that rocking movements preceded the onset of head banging. In 65% of the children head banging had begun before the first birthday, and in 2% it had started in the first 3 months of life. De Lissovoy (1962) noted that the most common positions for head banging were "on all fours" and sitting. All the mothers in this second study reported rhythmic behavior (such as body or head rolling, rocking, or swaying in highchair) prior to the onset of head banging. In the majority of cases, head banging occurred primarily at bedtime—75% of the children showed head banging every night or almost every night—but it also occurred upon waking and during the day. There was no significant association between the tempo of head banging and the sleeping heartbeat, but the correlation tended to show the heartbeat acting as a pacemaker. The tempo was also unrelated to the age of the children, and so De Lissovoy ruled out development of motor control as a factor affecting the tempo. In a follow-up study of these children, they were described as active, prone to rhythmic activity in response to music, and excellent in motor coordination.

---

[3] More girls were self-mutilators than boys, and Green advanced three possible explanations for this: a tendency for parents to discourage overt aggression in girls, the greater ability of male schizophrenic children to express their aggression in fantasy and play, and the fact that the female schizophrenic children came from more disturbed homes than the males.

[4] Recently, investigators have identified a sex-linked, recessive inborn error in purine metabolism that has been labeled the Lesch-Nyhan syndrome. The syndrome is characterized by hyperuricemia, hyperuricosuria, stunting of growth, retardation of mental and motor development, choreoathetosis, generalized spasticity, and compulsive self-mutilation of the tongue, lips, and fingers, predominantly by biting (Berman, Balis, & Dancis, 1969).

Kravitz et al. (1960) noted that body rocking also occurred in 67% of the head bangers in their sample. Temper tantrums were absent in 68% of the children, and crying while head banging was absent in 91%. There was an excess of firstborns over second borns (in contrast to De Lissovoy's finding), and head banging occurred in 20% of the siblings of the head bangers (higher than the reported incidence of 3.6%). Kravitz et al. suggested that hear banging starts when the teeth begin to appear, a factor that may partially account for its development. For many of the measures in this study (amount of separation, number of siblings, adequacy of mother–child relationship) Kravitz et al. did not employ a control group, and so some of their results are difficult to interpret.

Levy and Patrick (1928) claimed to find an association between the occurrence of breath holding and head banging in children and fainting and periodic headaches in their mothers. Although the association between breath holding in the child and fainting in the mother reached statistical significance, the other associations did not. Levy and Patrick did not attempt to account for their findings.

If head banging continues into adolescence, investigators view it as an abnormal behavior. Head banging is found in both neurotic and psychotic children. However, authors of case reports feel that the older the head banger, the more likely he is to be psychotic. Head banging rarely develops in older children or adults without a history of it in infancy.

Fitzherbert (1950) argued that the habit of head banging develops from the child's attempt to recapture the tempo of the mother's heartbeat that was experienced while being fed. She hypothesized that head banging should be rarer in breastfed children because they were held close to the mother's heart for only one-half of the feeding time.

Silberstein, Blackman, and Mandell (1966) noted that complete maternal deprivation leads to children becoming apathetic and depressed (Spitz, 1945), and they suggested that partial deprivation (rather than complete deprivation) causes head banging. They argued that the partially deprived child tries to provide immediate gratification of stimulation not provided by the mother. They also argued that the child tries to recreate the rocking movements made by the mother when she soothes him to recreate the intrauterine rocking he experienced when she walked prior to his birth.

Levy (1944) saw the origin of head banging (and other rhythmic movements) in sensory deprivation. He felt that infants subjected to movement restraint, monotony, or a poverty of play material would resort to rhythmic movements. Lourie (1949) felt that the rhythmic movements facilitated growth and development in the infant but that the movements acquired secondary value as a result of how the child's caretakers responded to the movements and how they treated the child.

De Lissovoy (1963) found a possible somatic basis for the origin of head banging. Children who were head bangers were significantly more likely to have middle-ear disease than children were were not.[5] However, middle-ear disease was present in only a minority of head bangers.

---

[5] A somatic basis underlying the origin of self-mutilating behavior does not appear to have general validity. Occasionally, cases of self-mutilation can be identified where there exists some localized sensation, such as an itch, that provides the initial somatic basis for excoriation (e.g., Mittelmann, 1933). The initial impetus for a self-mutilating behavior can also arise through chance, for example, through accidentally banging one's head. However, the adoption of this self-mutilating behavior must depend upon psycho-

### *Trichotillomania (Hair Pulling)*

Over the years, many cases of trichotillomania have been reported, and a systematic study of these cases indicates that the majority of them were females, adults, and psychotic. Greenberg and Sarner (1965) (without presenting any data from a control group) presented data on 19 individuals whom they had treated. Their typical patient was female, adolescent or preadolescent, from average to superior intelligence, white, Jewish, middle-class, whose hair pulling had begun at age 11–16, and who was diagnosed as either a "borderline" state or psychoneurotic (obsessive-compulsive). Comparison of Greenberg and Sarner's data with a typical case report indicates the possible bias that can result from taking isolated cases to be typical instances of the phenomenon. Of Greenberg and Sarner's patients, 10% admitted eating the hair, and a further 16% were observed eating it, although they denied doing so. Greenberg and Sarner described the mothers of hair pullers as ambivalent, restricting, critical, hostile, intolerant of aggression, highly competitive, overprotective and possessive, domineering, alternately acting infantile and parental, having strong unsatisfied dependency needs, and often being threatened by latent homosexual impulses. The fathers were seen as helpless, ineffectual, dominated and manipulated by the wife, aloof, withdrawn, passive-aggressive, and sometimes pseudoaffective socially. The mother–daughter relationship was very close and symbiotic, with a great deal of ambivalence, hostility, and fear of separation.

Delgado and Mannino (1969) also described the parents and their relation to the child. They felt that the child's needs were met very early in his life but that prior to the onset of his hair pulling, the mother withdrew and became only intermittently available as a love object. The frustration tolerance and the threshold for anxiety were low in the children. They were rarely able to express their aggressive feelings outwardly. Much has been made of the sexual overtones of hair pulling. The hair can be seen as a phallic symbol, and hair removal as symbolic castration. Case reports, especially those from psychoanalysts, tend to emphasize the sexual conflicts of the patients. Although the studies here have painted a reasonably complete picture of the family milieu in which hair pulling occurs, there have not been studies in which hair pullers are compared with other kinds of disturbed people.

### *Wrist Cutting*

Wrist cutting is quite common in psychiatric wards today, and there is agreement on its characteristics.[6] For example, Graff and Mallin (1967) described the typical wrist cutter:

> an attractive, intelligent, unmarried young woman, who is either promiscuous or overly afraid of sex, easily addicted, and unable to relate successfully to others. She is an older sibling with a cold, domineering mother and a withdrawn, passive hypercritical father. She slashes her wrists indiscriminately and repeatedly at the slightest provoction, but she does not commit suicide. She feels relief with the commission of her act [Graff & Mallin, 1967, p. 38].

Some wrist cutters can give reasons for the behavior. They talk of their pleasure and relief, their anger, or their depression. Offer and Barlow (1960) noted some other reasons: getting attention, the need to be loved and cared for, attempts to control aggression, tension reduction, and gaining prestige among the social group in the ward. The role of the social group in "epidemics"

---

logical factors, and an explanation solely in terms of the initial accidental discovery is inadequate. Furthermore, it is possible to discover a somatic or accidental initial stimulus for self-mutilating behavior only in rare cases. Indeed, some authors (e.g., Seitz, 1951) feel that the self-mutilating behavior occasionally is responsible for the somatic sensation rather than vice versa. Similarly, an explanation of the origin of self-mutilating behavior based on the fact that some people are indifferent to pain (e.g., Neil, 1958) is inadequate. In many cases, the production of pain is a crucial component of the behavior.

[6] Pao (1969) distinguished two kinds of wrist cutting: coarse and delicate. Coarse cutting was usually more serious and life threatening and performed by men over 35 years of age. Delicate cutting was nonserious and carried out by girls aged about 16 to 24.

of wrist cutting among patients in psychiatric wards has often been stressed (Matthews, 1968). There does not appear to be an association between the severity of the act and the degree of psychopathology (Goldwyn, Cahill, & Grunebaum, 1967).

Graff and Mallin (1967) noted that the wrist cutter demands and appears to benefit from physical contact with the therapist. The females reported that by their promiscuous sexual behavior, they were often seeking to be held and cuddled. Graff and Mallin felt that maternal deprivation probably played a part in the development of wrist cutting. The mothers of wrist cutters were described as cold, overpowering, hypercritical, hostile, and rejecting. This maternal rejection may be heightened by the birth of younger siblings. The wrist cutting serves to get attention and therapeutic help from others, to attack and punish the mother (by making her worried), and to provide the self-stimulation needed for relief.

Apart from describing limited samples of patients, the studies on wrist-cutting provided virtually no systematic, reliable information about wrist cutters. The studies never included a comparison group of noncutters, and the dependent variables were usually assessed by means of clinical judgments. Thus, the studies constitute hypothesis-generating research rather than tests of these hypotheses.

### *Self-Castration*

Blacker and Wong (1963) reported four cases of self-castration and reviewed other reported cases from the literature. (No case has ever been reported in the literature of a female mutilating her genitals.) In general, the patient was born to a mother aged 34–45, who was overcontrolling. The father was often absent from the home. There was a strong feminine identification and an overinvolvement with the mother. The men studied ranged in age from 29 to 66. Although sibling position does not appear to be associated with the phenomenon, Blacker and Wong argued for an excess of only children or younger children. Blacker and Wong also noted (*a*) severly impoverished childhood experiences, (*b*) long and intense sexual confusion, (*c*) submissive masochistic relationships with women, (*d*) depression relieved by genital mutilation, (*e*) strong feminine identification, and (*f*) repudiation of the genital organs. After castration, the men were usually affectless or cheerful. The castration appeared to resolve a long-standing conflict, and usually there had been a long history of viewing the genitals and sexual behavior as a source of trouble. The reports emphasized the psychosis of these men (especially the schizophrenic nature of their condition) and yet knowledge that the men had castrated themselves may have influenced the diagnosis. Mitchell (1968) described patients with self-inserted objects in the urethral passages and noted the similarity of these men to the self-castrators.

## SELF-MUTILATION IN ANIMALS

There are several reports of self-mutilating behavior in animals. Head banging has been noted in disturbed monkeys (Silberstein et al., 1966) and flesh biting has been observed in a poorly tamed rhesus monkey separated from his mate (Tinklepaugh, 1928). Self-mutilating behavior has also been observed during bursts of sham rage in decerebrate cats (Hendrick, 1950) and in monkeys with bilateral temporal lobe lesions (Kluver & Bucy, 1939).

Harlow (1962) raised monkeys in isolation with surrogate mothers and noted that, as adults, the monkeys rocked. When approached by humans (and occasionally when alone), they would show frenzies of rage and occasionally tear at themselves. A study by Mason and Sponholz (1963) was more systematic. They raised two monkeys in complete isolation and another two with visual and auditory experience of other monkeys (but no tactile contact) for 15 months. When the two isolated monkeys were placed together, they rarely made contact with each other but rocked and banged their heads quite often. When the two restricted monkeys were placed together, they frequently interacted. Occasionally, the male monkey of this pair would get excited, dance about his cage, pull his hair, and bite himself. Thus, the severity of isolation appeared to affect the choice of self-mutilating behavior.

There have been two reports of self-mutilation induced in rodents by chemical agents. Genovese, Napoli, and Bolego-Zonta (1969) injected pemoline gastrically and noted that within 6–8 hours, rats and mice showed amphetaminelike hypermotility and self-mutilating behavior, making lesions on the thorax and abdomen and frequently amputating their paws and tails. The experimenters found no effects from hunger or crowding. Peters (1967) injected caffeine gastrically into rats and found that they became very aggressive toward objects and mutilated themselves. The rats remained nonaggressive toward the experimenter, however. The effects were stronger in male rats, hungry rats, and isolated rats. Aggregation allowed fighting, and this reduced the incidence of self-mutilating behavior. Peters noted that the rats appeared to be especially sensitive to pain. There have been no analogous observations in humans. McManamy and Schube (1936), for example, described a case of caffeine-induced psychosis in a young girl, but self-mutilation was never observed.

## DISCUSSION

### *Self-Mutilation and Suicide*

Students of suicide have recently been debating whether attempts at suicide should really be considered as suicidal acts. Some suicidologists now prefer the terms *self-poisoners* and *self-injurers* to *attempted suicides* (e.g., Kessel, 1966). Others take the viewpoint that there is a basic similarity between attempted suicide (where the individual survives) and completed suicide (where the individual dies) (e.g., Lester, 1970a). Since attempted suicides survive and are more easily studied, most suicidologists have worked with them rather than with completed suicides. The idea that one may learn about completed suicides by studying other behavior has not extended far beyond the study of attempted suicides. One exception is Tabachnick (1970), who has begun to plan an investigation of those dying in car accidents in comparison to those dying from suicide.

There is seldom intent to die in the action of the self-mutilator. The individual may cause himself some pain, but, except in rare cases, he runs no risk of dying. Kessel (1966), in his study of attempted suicides, argued that many self-poisoners do not intend to die. By selection of the amount of poison consumed, the circumstances in which it is taken, and the provisions for rescue, the self-poisoner can ensure survival. Both nonlethal suicide attempts and acts of self-mutilation may be classified as *counterintentioned* suicidal acts, to use Shneidman's (1964) term in which self-destruction is threatened but not carried out.[7] Kessel argued that the self-poisoner tries to effect a temporary escape from a situation that is intolerable and also tries to manipulate the situation and the people involved by means of the act. Clearly, the secondary gain and manipulative aspects of the suicide attempt are similar to the secondary gain and manipulative aspects of acts of self-mutilation.

In addition to these similarities, there are some parallels in the research findings for suicidal behavior and self-mutilating behavior.

1. In suicidal behavior, men tend to use active methods (such as shooting and hanging) while women use passive methods (such as gas and pills) (Lester, 1969a). In self-mutilating behavior, wrist cutting and hair pulling are found more in women, whereas head banging and genital mutilation are found more in men.

2. The role of parental deprivation seems to be important in both kinds of acts. For suicidal behavior, investigators have found antecedent childhood experiences of complete parental deprivation through divorce or death to be common; whereas for self-mutilating behavior, clinical data indicate partial loss through emotional distancing and inconsistent maternal warmth to be common.

3. The role of anger, and difficulties in expressing aggression, have been noted in both behaviors. Suicide is often conceived as aggression turned inward onto the self, and, though this conception has not stood up well to empirical testing (Lester, 1968), it makes clinical sense.[8] When anger can be acknowl-

---

[7] Shneidman's other types of suicidal behavior are *intentioned*, *subintentioned*, and *unintentioned*.

[8] Rather than the suicidal person being less outwardly aggressive than the nonsuicidal person, Lester and Lester (1971) have suggested that the difference between the two groups lies in their control over aggressive impulses. Poor control may lead to

MITTELMANN, B. Allergic pruritis—neurotic excoriations. *Journal of Allergy,* 1933, **4**, 141–145.

MONTAGU, M. F. A. Ritual mutilation among primitive people. *Ciba Symposia,* 1946, **8**, 421–436.

NEIL, J. F. Self-mutilation of tongue. *Journal of Laryngology and Otology,* 1958, **72**, 947–950.

OFFER, D., & BARLOW, P. Adolescent and young adult self-mutilation. *Archives of General Psychiatry,* 1960, **3**, 194–204.

OLSEN, J. Self-inflicted injuries and false information to the police. *Medicine, Science, and the Law,* 1967, **7**, 22–25.

PANTON, J. H. The identification of predispositional factors in self-mutilation with a state prison population. *Journal of Clinical Psychology,* 1962, **18**, 63–67.

PAO, P. The syndrome of delicate self-cutting. *British Journal of Medical Psychology,* 1969, **42**, 195–206.

PETERS, J. M. Caffeine induced hemorrhagic automutilation. *Archives Internationales de Pharmacodynamie et de Thérapie,* 1967, **169**, 139–146.

PHILLIPS, R. H., & MUZAFFER, A. Some aspects of self-mutilation in the general population of a large psychiatric hospital. *Psychiatric Quarterly,* 1961, **35**, 421–423.

RESSMAN, A. C., & BUTTERWORTH, T. Localized acquired hypertrichosis. *Archives of Dermatology and Syphillis,* 1952, **65**, 418–423.

ROPER-HALL, M. J. Self-inflicted conjunctivitis. *British Journal of Ophthalmology,* 1950, **34**, 119–120.

SCHECHTER, D. C. Breast mutilation in the Amazons. *Surgery,* 1962, **51**, 554–560.

SEITZ, P. F. D. Psychocutaneous aspects of persistent pruritis and excessive excoriation. *Archives of Dermatology and Syphillis,* 1951, **64**, 136–141.

SHNEIDMAN, E. S. Suicide, sleep, and death. *Journal of Consulting Psychology,* 1964, **28**, 95–106.

SHODELL, M. J., & REITER, H. H. Self-mutilative behavior in verbal and non-verbal schizophrenic children. *Archives of General Psychiatry,* 1968, **19**, 453–455.

SILBERSTEIN, R., BLACKMAN, A., & MANDELL, W. Autoerotic headbanging. *Journal of the American Academy of Child Psychiatry,* 1966, **5**, 235–243.

SPITZ, R. Hospitalism. *Psychoanalytic Study of the Child,* 1945, **1**, 53–74.

STENBACK, A., & BLUMENTHAL, M. Relationship of alcoholism, hypochondria, and attempted suicide. *Acta Psychiatrica Scandanavia,* 1964, **40**, 133–140.

TABACHNICK, N. A theoretical approach to "accident" research. *Bulletin of Suicidology,* 1970, **6**, 18–23.

TINKLEPAUGH, O. L. The self-mutilation of a male macacus rhesus monkey. *Journal of Mammalology,* 1928, **9**, 293–300.

WECHSLER, D. The incidence and significance of finger-nail biting in children. *Psychoanalytic Review,* 1931, **18**, 201–209.

(Received October 22, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 2, 129–141

# INFLUENCE OF ECONOMIC AND POLITICAL IDEOLOGIES ON THE DEVELOPMENT OF DEVELOPMENTAL PSYCHOLOGY[1]

KLAUS F. RIEGEL[2]

*University of Michigan*

Two trends in developmental psychology are delineated: one representing the "capitalistic" orientation of the Anglo-American countries and elaborated within the Hobbes-Darwin-Galton-Hall tradition; the other representing the "mercantilistic-socialistic" orientation of the European continent and elaborated within the Rousseau-Pestalozzi-Spranger-Piaget tradition. The former implies a model of continuous growth in which all individuals are evaluated against single standards, such as the white achievement-oriented intelligence test; the latter implies qualitative growth models and encourages multigenerational and multicultural standards. In the first two parts of this article, the two major trends are described. In the third part, an integration of these viewpoints is attempted by emphasizing a dialectic interaction model and by relying on Rubinstejn's notion of "constitutive relationism." The dependency of growth upon sociocultural contingencies is demonstrated through studies on the recall of historical events; the dependency of growth upon biological contingencies is demonstrated through theories from cellular biology. Finally, the dynamic character of the interaction processes, that is, their changes with individual and historical time, is discussed in reference to modern developmental research designs.

The recent unrest in our society has made us once more aware of the intimate connections between economic and political ideologies and the development of sciences. This article does not intend to draw upon the well-documented preferences of most governments for defense-related rather than socially significant spending or upon the political frustrations and problematic decisions that modern scientists have to face, but rather upon the general economic and cultural conditions of societies that provide a necessary basis for the growth of sciences. This dependency can be persuasively documented for developmental psychology, which exploits two traditions, identifiable by certain philosophical, educational, economic, social, and political orientations. For convenience, these are labeled the "capitalistic" and "mercantilistic-socialistic" traditions. The first has been dominant in the Anglo-American world; the second has prevailed in continental Europe. These traditions are discussed in the first two parts of this article. In the third part, the author attempts to synthesize these two trends in what might be called an interactional or relational interpretation of developmental processes. The third part, therefore, suggests "what we must do"[3] in developmental psychology.

## THE "CAPITALISTIC" ORIENTATION

Most textbooks on the history of psychology, if they are treating the topic of individual differences and development at all, pay much, if not exclusive, tribute to Darwin's (1859) theory of the "Origin of Species." Darwin (1809–1882) revived, on the basis of rich naturalistic observations, some ideas that are familiar to the British philosophical tradition and that are expressed most distinctly in Hobbes' (1588–1679) analysis of the development of social organizations. At the beginning, Hobbes (1669) argued in his "Elementa Philosophica de Cive" each individual is engaged in a struggle against everyone else. Being motivated by egoism, only sufficient damage makes man realize that his security and future is guaranteed only if some cooperation and social contract is achieved. Such arrangements ultimately lead to the establishment of reasonable social orders, especially through the development of property rights.

---

[1] This paper was given as a special lecture on "Language and Cognitive Development," also attended by members of the Black Action Movement prior to their successful strike for greater academic and social participation at the University of Michigan.

[2] The author is greatly indebted to the participants for their engaged discussions and encouragements. Special thanks belong to Inge Ahammer, Clinton Fink, John Meacham, Herbert Odom, Carole Quarterman, Ruth Riegel, George Rosenwald, Karen Spielman, Elliot Stern, and Ronald Wacker.

Requests for reprints should be sent to Klaus Riegel, Department of Psychology, University of Michigan, Ann Arbor, Michigan 48104.

[3] This expression is borrowed from the title of the challenging article by John Platt (1969), which in turn seems to have been taken from Lenin's pamphlet (1929), which in its English translation reads "What is to be done."

1. Darwin, with his notions of the "struggle for survival" and the "survival of the fittest," proposed an open-ended developmental interpretation quite similar to the social philosophy of Hobbes. For Darwin, development is a process of continuous competition and selection whose direction and goals are represented by the "successful survivors" here and now. When translated into the matrix of behavioral and social sciences, the "successful survivor" becomes the white, middle-class adult most likely engaged in manufacturing or business enterprises.[4] His success was not so much seen as being due to his generative capacities but rather to the fit between his activities and the demands of the society that he represented.

From this point of view, deviant persons or nonstandard groups attain negative attributes only: children are regarded as incomplete adults; old persons as deficient; criminals, mental defectives, colonial subjects, and nonwhites as far below the rank of white middle-class adults. It cannot be denied that many of the privileged, especially in Britain, vigorously engaged in humanitarian efforts to better the "deplorable" lot of these persons, but in all these activities they never abandoned their basic ideological and evaluative schema. Instead, they reinforced it through these efforts in much the same way as the role of the beggar is reinforced by giving him alms. Occasionally, they pitied the fate of their own making as the "white man's burden."

Darwin's theory of evolution was most readily accepted in England. France was most resistant, and Germany took an intermediate position (Norderskiöld, 1928). These national differences can be explained on the basis of economic and cultural conditions. The Englishmen with their cherished tradition for hunting and breeding—but especially the Englishmen as entrepreneurs, traders, and manufacturers—should be more than superficially susceptible to interpretations in which success emerges through restless competition. Of course, the idea of an apelike ancestry shocked the more conservative circles, but such considerations were easily outweighed by the relevance and similarity of these interpretations with the politico-economic ideals of a free, competitive trade system.

While the strong emphasis upon competition placed the young, the old, the deviant, and, generally, the "different" persons into inferior positions, the positive identification of the successful adult was, nevertheless, far from a simple manner. By posing an open-ended ideal, such a prototype became more a topic of wishful thinking than an observable phenomenon. Dependent upon personal preferences and values, the ideal may represent the imaginative manufacturer, the aggressive businessman, the clever politician, or the ruthless general; dependent upon stereotypes and illusions, it represents adventurers, movie stars, princes, and charlatans. In emphasizing prototypes rather than real people, it mattered little that the private affairs and mental states of such ideal persons (consensual heroes) might be pitiful or even more deplorable than those of the average and presumably less successful and happy individuals. The goal for every member of the society remained to strive toward these abstracted ideals and if unsuccessful to be regarded with lesser respect.

2. One of the earliest proponents of the theory of evolution and of its translation into the behavioral and social sciences was Francis Galton (1822–1911). Under the strong influence of his older sister, and endowed with sufficient wealth, Galton developed into a brilliant superman-scientist with a feminine

---

[4] This translation was promoted by Spencer (1897) (1820–1903) and Pearson (1904) (1857–1936) in Britain, but especially by Sumner (1963) (1840–1910) in the United States.

sensitivity, as described in Nietzsche's writings and also as demonstrated through Nietzsche's life. Aside from extensive travels and explorations, Galton conducted numerous innovative studies on crossbreeding, inheritance of traits, and individual differences. He invented and applied many methods, such as the free-word-association and psychomotor tests, as well as procedures for their analysis, such as the regression technique. Most important for our present purpose are Galton's "Hereditary Genius" (1869) and "Inquiries into Human Faculty and its Development" (1883).

Emphasizing a biogenetic basis, Galton, in the first book, traced the lineage of famous persons in order to show the hereditary nature of superior abilities. In his conclusion he did not hesitate to argue for selective inbreeding of gifted persons in order to improve the human race. This perspective led him to establish the field of eugenics, which he justified as follows:

> It may seem monstrous that the weak should be crowded out by the strong, but it is still more monstrous that the races best fitted to play their part on the stage of life, should be crowded out by the incompetent, the ailing, and the desponding [1869, p. 343].

Supported by the like-minded Count Gobineau (1816–1882; reference, 1884) and generalized to cultural-historical comparisons by Houston Stewart Chamberlain (1855–1927; reference, 1909) (the uncle of the late British prime minister Neville Chamberlain and son-in-law of Richard Wagner), these ideas, when merged with continental European ideologies and applied with political fanaticism and military expediency, led directly to the holocaust of Belsen, Buchenwald, and Theresienstadt.

In his second book, "Inquiries into Human Faculty and its Development," Galton (1883) introduced impressive biological, psychological, and anthropological records among which his psychometric measures—reaction time, strength of grip, coordination speed, etc.—across the full life span still are some of the most accurate and extensive data available (see Koga & Morant, 1923; Ruger & Stoessiger, 1927). Other reports were more dramatic but less serious. Thus, during his numerous journeys in England, he counted the number of beautiful women seen per unit time and in his "beauty maps" reported that London had most and Aberdeen fewest of them. Nevertheless, his extensive recordings became the prototype of a developmental research strategy which—at least on the North American continent—was to dominate this field for many decades to come. Theoretically more sophisticated approaches, as well as experimental methodologies, have only recently begun to play a significant role in this branch of behavioral science. In comparison, intensive case studies of infants and young children became more characteristic for developmental research in continental Europe than in England (see Stern & Stern, 1907), even though Darwin (1877) himself contributed a "Biographical Sketch of an Infant" during the later part of the nineteenth century.

Galton's thinking was dominated by the biographical and ecological interpretations of his famous cousin. These interpretations were based on descriptive evidence and, from our present point of view, resemble scientific speculations rather than scientific theories. Even though progress was soon to be made, for instance, through the rediscovery and extension of Mendel's laws of crossbreeding by Correns, de Vries, and Tschermak, it took several decades until an experimental basis for genetic studies and appropriate methodologies for their analysis were developed. Undoubtedly, Galton was one of the early contributors to this development.

3. Similar though less credit has to be given to the American followers of Darwin's theory, especially G. Stanley Hall (1846–1924). While his continental European colleagues, including such outspoken promoters of the theory of evolution as Ernst Haeckel (1834–1919), engaged in experimental genetics and comparative anatomy and physiology, Hall tended to emphasize far-reaching generalizations rather than tedious research details. He strongly supported the theory of pangenesis with its notion that "ontogeny recapitulates phylogeny." In these interpretations Hall deviated from Galton by placing greater emphasis on environmental factors. While Galton, for instance, supported the no-

tion of *gemmules* as carriers of hereditary information, arguing that they were freely circulating through the body, Hall made some concessions toward Lamarck's (1744–1829) notions of the inheritance of acquired traits and of changes actively accomplished by organisms. In this regard, surprisingly, he found himself in agreement with Darwin himself, who also took a rather liberal view when dealing with questions of human development. Galton was indeed the extremist, arguing for the inheritance of mental traits, whereas Hall's interpretation represented a compromise between hereditarian and environmentalistic views. Recently, these differences have been cogently discussed by McCullers (1969) and Charles (1970).

Up to adolescence, Hall argued, development is biologically determined; thus, one has to allow "the fundamental traits of savagery their fling till twelve [Hall, 1904, p. X]." During adolescence and early maturity, however, genetic changes could be effected, thus, society has a heavy burden. Moreover, as expressed by Hyatt (1838–1902), an American naturalist:

> all modifications and variations in progressive series tend to appear first in the adolescent or adult stages of growth, and then to be inherited in successive descendents at earlier and earlier stages according to the law of acceleration, until they either become embryonic, or are crowded out of the organizations, and replaced in the development of characteristics of later origin [1890, p. IX].

Hall's compromise in the United States and Pearson's psychometric zeal in England initiated the long-lasting struggle between supporters of hereditary interpretations and the environmentalists. This controversy still burdens heavily our conceptualization of development as the recent excitement about Jensen's (1969) review and the reply by the Society for the Psychological Study of Social Issues (1969) have revealed. As is discussed in the next section, the continental European, developmental psychology is free of such an extrapsychological issue by dealing with psychological development in a nonreductionistic manner; that is, by trying to explain psychological development on a psychological rather than on a biogenetic or sociocultural basis.

Hall's compromise began to shift the emphasis from the competition between genetically determined to that between socially determined factors. This shift, much more than Galton's biogenetic interpretations and his notion of an applied field of eugenics, was to influence the developmental views in the major capitalistic country of today. Hall's students, especially Terman (1887–1956) and Gesell (1880–1961), continued within the empiricistic orientation of Galton, however, and the following decades saw a burst of descriptive information on child development in the United States. All these studies contributed to establishing trends and standards with which the performance of the young and old, the abnormal and the deprived, could be compared. As long as the middle-class, white adult remained the sole standard-ideal, the outcome of all these evaluations was bound to remain negative.

## THE "MERCANTILISTIC-SOCIALISTIC" ORIENTATION

Since the economic and social structure of a nation is strongly determined by geographical factors, it is not surprising that England emerged as a strong power at the seas and as a nation of entrepreneurs and merchants successfully trading their manufactured goods across the continents. Even though a monarchy, the government was not suppressive and the taxation was not severe enough to hamper seriously the industrial and financial developments. Colonial exploitations served to foster economic growth further, which reached its culmination during the late nineteenth century (Beard, 1962). These conditions led to the emergence of a specific developmental science, representing the capitalistic mode of thinking.

In comparison to England, quite different economic and social conditions prevailed on the European continent, especially in France and in major parts of Germany and Italy. These conditions brought about a distinctly different type of developmental science which might be called its mercantilistic-socialistic form.

1. All three countries, but especially France and Germany, remained land powers and were

dominated by warrior aristocracies that controlled large land holdings. While, in general, the industrial and economic progress was delayed and slower than in England, the maintenance of large armies and courts made the development of manufacturing necessary. These developments were usually initiated and supervised by the state, with the occasional exception of military production during times of increased demands. They were either manipulated by the rulers themselves, such as in Prussia, or by powerful advisors such as Jacques Coeur (1395–1456), Maximilien Sully (1560–1641), Baptiste Colbert (1619–1683), and Jacques Necker (1732–1804) in France. Whenever exceptionally successful, these entrepreneurs were ultimately suppressed, however, and the French Houguenots (especially after the cancellation of the Edict of Nantes in 1685) were known to have fled in large numbers to neighboring countries where they contributed much to the development of trade and industry, such as in England, in the city ports of the North Sea, and also as far away as in Russia.

The increased amount of manufactured goods and the increased trade within the continental European countries led, nevertheless, to the rapid expansion of a third, middle class of citizens which, though better off than the laborers, servants, and farm workers, was less privileged than the ruling aristocrats. Eventually, the incongruency between their social significance and their social privileges led to the French Revolution.

From the perspective of modern Western ideology, the prevalence of distinct classes with graded duties and countergraded privileges seems less than desirable. On the positive side, however, these conditions helped to strengthen an awareness and sensitivity toward group and age differences. In philosophy it produced the conservative attempts to justify the existing social orders, for instance, as implied in Leibniz' (1646–1716) Theodicee and in Hegel's (1770–1831) stages of developmental progression. By disregarding ceaseless competition between (but not within) classes, the continental European conditions generated, on the other hand, a social and educational philosophy appreciative of multicultural and multigenerational differences. The home base for this movement lies at Geneva, the precise intersection of the three major nations involved.

2. Influenced by Leibniz and supported by German representatives of philosophical Enlightenment, such as Mendelsohn (1729–1786), Lessing (1729–1781), and Herder (1744–1803), the first key figure in this development is Rousseau (1712–1778). In Rousseau's (1750) treatise on the future of science, but especially in his (1762) book "Emile," he outlined an educational philosophy in which the child is set apart from the adult world of comparison and is educated and evaluated in view of the standards of his peer group alone. In support of a romantic ideal of a "noble savage," and in sharp contrast to Hobbes, men—and in particular primitive men and children—are considered as basically good. The adult world and civilization resemble a restrictive cloak that covers and spoils the natural beauty of the mind and body. Men are basically equal. Civilization creates individual and group differences that make some rich and powerful and others poor and feeble. Rousseau's educational philosophy aimed toward preserving as long as possible the natural state of the child and tried to reaffirm the equality of men.

In a short period of time, Rousseau's educational and social philosophy began to influence the emerging educational institutions on the European continent. Although public education had become obligatory during the eighteenth century, these attempts mainly served the purpose of advancing the production potential of these states. Little thought was given to educational strategy, and since the teaching was often in the hands of retired army sergeants, drill and punishment were the predominant means of training. With the rapid spread of the new educational philosophy, especially through Pestalozzi (1746–1827), attention was given to educational approaches appropriate for the child, and the profession of specially trained teachers began to emerge in the more advanced European countries.

While this movement, initiated by Rousseau, can be considered as a revolution in support of the child, other upheavals originated from within the youth groups themselves. The early student organizations in the German states represent the first case in point. Originally conceived as paramilitary training groups against Napoleonic suppression, these organizations began to carve out distinct roles and privileges for students and teachers that were to influence the academic community for generations to come and are only now challenged by the radical student movements in the European universities. (Unfortunately, these old but originally progressive fraternities became so rigid during their long history that major social and political changes began to bypass them without traces, and they have since attained the image of reactionary relics.)

After the educational breakthroughs for the pupils of elementary schools and by the students at the universities, the next movement was in support of preschool education and is especially connected with the names of Fröbel (1782–1852) and Montessori (1870–1952). The former, in close affinity to Pestalozzi, introduced the child-centered approach in "kindergarten," a term coined by him. Maria Montessori, giving stronger emphasis to an achievement orientation, emphasized, nevertheless, a child-initiated and self-paced type of educational strategy. Both charted a special role for and educational approaches to the preschool child, the former most liberal and laissez-faire, the latter more rationalistic and goal directed.

The next major movement was again youth initiated and influenced several better-known educators and developmental psychologists who, during their early years, actively participated. This movement, called the "youth movement," was concerned with the liberation of adolescents from the establishments of the Victorian adult culture. It sought a "return to nature," and expressed itself in a preference for scouting, hiking, and camping. Unlike its counterpart in the capitalistic countries (which persisted much longer), it was dominated by the thought that youth should be led by youth, and, therefore, the leaders were exclusively chosen from within the groups themselves.

As a participant in his youth, Spranger (1882–1963) became a developmental psychologist who formulated his educational experiences within a phenomenological psychology of understanding and empathy. In Spranger's (1924) influential book on adolescence he claimed that the thoughts and culture of this group, as well as of their deviant (i.e., their so-called "delinquent") members, should be appreciated on the basis of sympathetic understanding and should not be evaluated in terms of the adult culture within which they exist uniquely.

3. Spranger's interpretations might stand for those of a good many other continental European psychologists who formed a group quite distinctly apart from their American and British contemporaries. The continental European movement originated, flourished, perpetuated, and spread from Geneva and reached its latest peak in the work of Piaget. This movement is characterized throughout by the notion of a stepwise, rather than continuous, progression in development. Each stage has to be evaluated in the framework of its own standards and, strictly speaking, is incompatible with any other level of behavior. While competition within stages is conceivable, it is not of great importance and is rejected as a driving force across stages.

Piaget has been criticized for his disregard of environmental, especially linguistic, factors and their impact upon cognitive development. Piaget's view is understandable within the developmental interpretations preferred among most continental European scientists (see Aebli, 1951). Language as a universal matrix for shaping the thoughts of the growing child is disregarded in favor of the conception of language as a succession of specialized sublanguages characteristic of different developmental stages. This distinction has been most clearly expressed by the Geneva school of linguistics, especially by de Saussure's (1904) contrast of *la langue* and *les paroles*.

The early controversy raised in Piaget's (1926) discussion of egocentric and socialized speech of children suggests, however, a secondary conceptual distinction within the European tradition. Piaget, agreeing with Hobbes rather than Rousseau, regarded the young child as self-centered and egoistic. Development changes the individual into a social being. This interpretation was radically challenged by Vigotsky (1962), who, relying on equally dramatic observations, argued that children are basically social. Experience induces a self-concern and awareness of the overt social, for example, linguistic actions. This progression, ultimately, leads to thought as an internalized form of speech. At the midway point, that is, around 6 or 7 years of age, the child produces self-centered speech that is not yet completely internalized and also lacks social and communicative features.

Vigotsky's (1894–1934) approach, which is congruent with that of other modern Russian psychologists, might be called the socialistic interpretation within the continental European tradition of developmental psychology. Piaget, on the other hand, represented its right branch and hardly deserves the same label. Most appropriately, this distinction might be compared with the neo-Hegelian philosophy during the middle of the nineteenth century which broke up into a conservative right branch and the better-known left branch represented by Strauss (1808–1874), Feuerbach (1804–1872), and others. This branch eventually led into the dialectic materialism of Marx (1818–1883), Engels (1820–1895), and Lenin (1870–1924).

## AN INTERACTIONAL OR RELATIONAL INTERPRETATION OF DEVELOPMENTAL PROCESSES

The influence of continental European conceptualization is rapidly growing among American scholars as well as within the American society. Such a reorientation seems most appropriate. It not only assigns appreciable places to individuals at various stages of development, and more humane roles to the aged and deviant, but also allows us to deal fairly with various subgroups and subcultures still striving to find their own identity. While the Anglo-American tradition has come to measure all individuals and groups along the same yardstick, the continental European educational and social philosophy promotes a diversity of standards and, thus, is sensitive to multigenerational and multicultural aspects.

Undoubtedly, the continental European social philosophy has serious shortcomings. In its extreme version it promotes the establishment of classes and castes. However, such implications might represent the lesser evil under conditions where the communication between generations and between incompletely realized subcultures is either lacking altogether or has been weakened to such an extent that no other means than revolution might induce the established groups to change their perspectives in order to make communication and mutual appreciation possible again.

But in spite of other conclusions that might have been drawn from these statements, we do not promote the adoption of either of the two views, because science and knowledge, as well as the society in general, can advance only if the divergent viewpoints are integrated at higher and more abstract levels. Behavioral scientists and developmental psychologists in particular have not succeeded in consolidating or synthesizing the two viewpoints presented. This failure has to be explained by a lack of educational, social, and political awareness on the part of the individual scientist, by his preference for mechanistic rather than dialectic considerations, and by his naive "fact-finding" orientation.[5] Recently these issues have been brought into the open in theoretical discussions (see Holzkamp, 1970), and by the promotion of interaction models.

1. Undoubtedly, the idea of writing a psychology in interaction terms is not an original one. Such attempts have been suggested by several continental European as well as Anglo-American scholars. For instance, von Uexküll's (1909) interaction paradigm is gaining much delayed though increasing attention from ecologists studying organisms in their natural environment. While von Uexküll (1864–1944) avoided any abstract formalism in his interpretations, Lewin (1890–1947), with due recognition of von

---

[5] Such a naive "fact-finding" orientation is clearly revealed by a quotation from Jensen in a recent report in *Life:* "I don't see why people should be disturbed by unequal representation of different groups in different occupations—or educationally, *if* it should be found that there are real differences [Neary, 1970].

Uexküll, tried to provide such an analysis (Lewin, 1936). Also Anglo-American psychologists have proposed interaction models, most notably Kantor (1959) in his interbehavioral psychology, without receiving appropriate recognition.

For several decades a similar fate seemed to have befallen Piaget until his contributions were rediscovered through the efforts of Berlyne (1957), Hunt (1961), and especially Flavell (1963). Piaget (1968) based his theorizing on the paradigm of a dialectic interdependence between the processes of accommodation (of the subject to the object) and assimilation (of the object to the subject) leading to continuous adaptations and readaptations. While Piaget is a strong case in point, the most explicit theory of interactions has been provided in Chomsky's transformational grammar. This interpretation of Chomsky's contribution has not been sufficiently recognized by psychologists.

Originally, Chomsky (1957) regarded linguistic transformations as devices that would make the production of utterances more efficient by deriving them from a limited set of kernel sentences rather than by deriving each sentence independently and anew. In his recent writings (1965), he has given a more central role to the concept of interaction, however, by proposing a distinction between a universal base structure and the surface structures of the utterances produced. By carrying this idea still further, one might disregard the surface structures as superficial and idiosyncratic, and the abstract base structure as intangible. What remains and what represents the essential features of language is the system of transformations.[6]

Such an interpretation of Chomsky's theory may seem one-sided. But, if nothing else, it makes us apprehend the insufficiencies of the traditional approaches to the problem of interaction. Looking at language again, we can see that these insufficiencies have been created by an undue emphasis on linguistic particles at the expense of contextual interpretations. Prior to Chomsky, language acquisition was seen as consisting, first, in the learning of elements, such as phonemes or words; and second, in the learning of sequential dependencies which, ultimately, were expected to explain the production of meaningful and complex utterances. In this writer's opinion (Riegel, 1970a, 1970b), such interpretations reflect nothing but the bias of the analytical scientist. We cannot seriously doubt that all linguistic as well as perceptual-cognitive experience is relational. We receive elements always in contexts, never in isolation. Words, for instance, are always related to other words or to the objects, events, or qualities that they denote. Only in the ivory towers of the psychological laboratories, with their focus upon nonsense syllables, or in the existentialistic quarters, with their interest in such statements as "a rose is a rose is a rose" (Gertrude Stein), are elements seemingly viewed in isolation. These exceptional considerations aside, relations are given immediately; elements as well as classes are derived or abstracted from relational information, and, in this way, interactions are intimately embedded in our experiences.

In modern philosophy, the idea of a relational interpretation originated with Hegel's dialectic idealism and reappeared, especially, in Lenin's writing. Rubinstejn, a Russian psychologist, has given it renewed expression in what he called "constitutive relationism according to which every phenomenon or thing is determined and constituted by its relation to all the other phenomena of reality [Payne, 1968, p. 99]." Rubinstejn[7] emphasized the material basis from which relations originate and through which they generate double interaction systems, namely, between psychic and external sociocultural conditions (representing the dialectic materialism of Soviet psychology emphasized after the First All-Union Conference on Human Behavior in 1930) and between psychic and internal biological conditions (representing the mechanistic materialism emphasized prior to this conference). Similarly, Rubinstejn (whose synthesis of these two movements gained acceptance after the Pavlov Conference of 1950) regarded the material world as having both an internal structure, for example, the molecular structure of a crystal, and an external structure that relates the crystal to other objects as well as to subjects. The observable results of this external interaction can be defined by the crystal's tangible qualities; for example, the crystal will appear hard and pointed to a particular observer. Finally, both interactions represent processes within individual and historical developments, which they both influence and by which they are both influenced. For a clarification of these issues, let us consider the following experimental demonstrations.

---

[6] Strong support for such an interpretation is given by Piaget (1970) who stated, "Transformations may be disengaged from the objects subject to such transformation and the group solely defined in terms of the set of transformations [Piaget, 1970, pp. 23–24].

[7] It is not a purpose of this article to give a comprehensive survey of Rubinstejn's theory. For obtaining such information the reader is urged to consult the recent book by Payne (1968).

2. If young adults are asked to recall the names of all the persons they had met during their lives, and if one plots the number of persons recalled against the years at which they were met for the first time, more names of persons who had been met either most recently or very early in life will appear than those who were met during the intermediate periods. These results (Riegel, 1972) raise the question of the extent to which the psychological recall scores reflect the conditions of the ever-expanding sociocultural environment and of the changes in motility that allow growing individuals to explore their social possibilities to an ever greater extent.

Thus, at least two components need to be considered in an interpretation that tries to explain the relations between psychological recall data and nonpsychological contingencies. On the other hand, the interaction with subject's growing motility has to be analyzed which, in turn, might depend upon internal maturational sequencing (e.g., upon "critical periods" of his learning to sit up, walk, speak, write, etc.). On the other hand, and as emphasized by Brunswik (1949) and Barker and Wright (1951), the interaction with shifts in external contingencies has to be studied. Much like the internal biological sequences, often these external sociocultural changes are beyond the control of the growing organism. For instance, the average individual is born into a family of a given size and structure, grows up in a given neighborhood and, because of educational regulations, enters different types of schools at predescribed points in time and with predescribed, increasing numbers of other children. The number of new persons met later in life will vary greatly between occupations; few changes might occur for decades, but a minimal amount of interaction will continue to exist at his home, church, business or job site; only when retirement arrives is his social and communicative potential suddenly disrupted.

A stronger case yet for the second type of interaction, that is, for the influence of sociocultural contingencies upon subjective experiences and recall, can be made by studying the free recall of historical figures. The process of recalling such persons (e.g., politicians, scientists, artists, etc.) differs from the recall of names of friends and acquaintances in that there is no zero point in time comparable to the birth of the recalling subject. Moreover, the progression is not as well charted as through the shifts to different school types, to different geographical locations, or to different job sites. Nevertheless, the results reveal quite similar tendencies. For the majority of American college students, history begins at a fixed point, namely, with George Washington, and, thereafter, progresses from catastrophe to catastrophe, from war to war, at which point names of historical figures accumulate far out of proportion to the steady increase with historical time in the size of the population.

An inspection of these data alerts us to question whether history "really" progresses by such "leaps and bounds" or whether these apocalyptic steps merely reflect the biases with which politicians, educators, and historians indoctrinate the minds of historical perceivers. The data make us question why we give so little attention to the slow but steady progress in economic and social welfare or to the advances in sciences and arts that might counteract the wavelike sequences of military catastrophies. Most important, our results indicate how tightly social conditions and political ideologies are connected with subjective experiences. But they also suggest that once we have become aware of this interaction, individuals through new efforts in their teaching and writing might change the perception of our social and intellectual environment and, thus, may change this environment itself for generations to come (Riegel, 1969).

These examples demonstrate three components of the interaction processes: the sociocultural contingencies (set by environmental conditions and regulations), the biological contingencies (set by biochemical and biophysical determinants), and the psychological processes (emerging into consciousness through interactions). In the present section, examples on the recall of past events were used to demonstrate, primarily, the interaction between sociocultural contingencies and psychological performances; the following section refers to some biological theories of growth and draws some inferences on the interaction between biological contingencies and psychological performances.

3. As an analysis of the sociocultural contingencies enables us to explain differences in the recall of past events, so the biological theories, to which we are referring, suggest possibilities for explaining growth. Traditionally, developmental sciences have restricted their inquiries to descriptions, occasionally couched in such fancy terminologies as phenotypical versus genotypical representations, epigenetic phenomena, zeitgeist, etc. On other occasions, developmental scientists have tried to obviate the problem of explanation by emphasizing that time itself can serve explanatory functions, a statement that implies nothing else than to say that, for instance, 4-year-olds have certain height because they are 4 years old.

Even though neither conclusive nor generally accepted, cellular biologists have proposed two interpretations that explain growth and do not merely describe it. The first has been called *waste theory*. Basic metabolic processes of cells lead to the production of waste products, such as lipofuscin, which cannot be completely removed. These products accumulate over time until they reach a critical level and produce a decrease in functioning. According to this theory, life itself implies growth and aging via the accumulation of metabolic waste. The waste theory represents an intrinsic theory of development: biological age is determined by the amount of waste accumulated by and within the organism.

The second theory has been called *mutation theory*. Mutations—which are generally deleterious for the organism—occur at random. The older a person is, the greater the number of mutations that must have occurred and the more likely that structural or functional defects will have resulted. Since the mutation rate depends on the amount of irradiation to which an organism has been subjected, some theorists regard the amount of radioactive fallout as a major determinant of development and aging. This formulation is important, not only because it suggests ways of manipulating developmental processes in the laboratory, but also because the determinant of age has become an extrinsic, nonbiological entity.

These two biological theories of growth represent at a molecular level the two types of interactions delineated by Rubinstejn. Our interpretations of the recall of persons represent molar counterparts. Undoubtedly, the gap between these two levels of interpretations cannot be bridged at the present time and—in this author's opinion—should not be bridged in a reductionistic manner. What is needed is a conceptual integration. Let us consider, for instance, changes in verbal reaction times during the life span. On the one hand, performance speed is dependent upon a well-provided repertoire of responses. According to available evidence and interpretations (Riegel, 1966, 1968), this repertoire increases with age (even though to an increasingly smaller extent). Subsequently, the speed of responding might increase. On the other hand, it is known and, eventually, might be explained on the basis of biological theories of growth, that there is a continuing deceleration with age in the speed of sensorimotor activities that will counteract the continuous improvements expected on the basis of the language model. The joint outcome of these processes is an increase in response speed up to earlier adulthood and a decline thereafter.

4. According to Rubinstejn, psychological processes are constituted by two interactions,

those with external sociocultural contingencies and those with internal biochemical conditions.[8] As an outcome of these interactions, the organism changes; as the organism changes, the interactions change. In other words, interactions interact with time. This intimate connection between psychological changes in the individual and the changes in intrinsic and extrinsic contingencies has only recently received appropriate emphasis in developmental studies. The present section draws attention to models of developmental analyses proposed by Schaie (1965), which seek to explicate the dynamic character of the interaction processes described.

In his analysis of the extensive testing records of military draftees during World War I and II, Tuddenham (1948) noted that the mean intelligence of the recruits of 1941 coincided with the 82nd percentile of those of 1917. Therefore a considerable improvement in intelligence was observed which (if both samples could be regarded as representative of the same population) could be due to the better economic conditions, education, health care, etc., of the younger generation (i.e., to changes in the external, sociocultural contingencies). If the subjects of 1917 were retested in 1941 (by which time they would have been over 40 years old), deficiencies of their retest performance in comparison to that of the 17-year-old draftees of 1941 could not necessarily be attributed to aging (as it is usually done in such cross-sectional comparisons), that is, to changes in the intrinsic, biological contingencies. Since the 40-year-old subjects of 1941 did not attain, when they were young, the level of the 17-year-old draftees of 1941, the results indicate equally well the enormous sociocultural changes, for instance, in educational and communicative conditions, that is, differences between cohorts (Riegel, 1972; Ryder, 1965).

While the traditional method of developmental psychology (the cross-sectional comparison), is insufficient for unconfounding the interaction of individual and historical changes, the same conclusion needs to be drawn for the longitudinal research methodology which (in the judgments of most developmental psychologists) ranks so much higher than cross-sectional comparisons. If, in 1941, we compared the retest results of the draftees of 1917 with their original performance, such a longitudinal comparison would have been accomplished. But we would have equally failed in determining whether any observed differences were due to the individual's intrinsic changes or to those in the extrinsic sociocultural conditions affecting the individual (in this case, between the first and second testing rather than during the time periods prior to a cross-sectional comparison of 1941). Most likely, both components will be effective in a confounded manner.

Schaie's (1965) and Baltes' (1968) research designs for developmental studies make possible the unconfounding of the interaction between the changes in psychological and in extrinsic, sociocultural conditions. Their success is due to the recognition of the third basic design embedded in such a set of data in addition to the cross-sectional and longitudinal arrangements. This design (called "time-lag" design) compares the performances of one age group, tested at different historical times; for instance, it compares (as did Tuddenham) the performances of 17-year-old draftees in 1917 and 1941. Since age is held constant, only sociocultural changes can become effective. In conjunction with the other two basic developmental designs it becomes possible, therefore, to estimate the relative contributions of these extrinsic changes upon psychological growth.

According to Schaie, similar explications are possible for the interaction between psychological growth and intrinsic biological changes. Such analyses would require a basic design analogous to the time-lag design except that sociocultural rather than the biological changes of individuals need to be held constant. Although it might be difficult to conceive of changes in individuals that occur in a sociocultural vacuum, attempts of raising children in complete isolation have been reported in the literature (see Brown, 1958), the first of which was attributed to Pharao Psammetichos

---

[8] The Russian reflexology represents another example of the second interaction process preferred in the discussion by Rubinstejn.

I, over 2,600 years ago (see Herodotus, 1931). Naturally, these quasi-scientific experiments did not resolve the problem satisfactorily, since the children remained affected by changes in the physical conditions such as the weather, diurnal and seasonal cycles, etc., which are part of the extrinsic contingencies. On the other hand, these attempts are not as exceptional as they might appear upon the first impression. This recognition leads us to a final conclusion, which has been implied throughout this article.

In spite of our extraordinary advances in science, technology, and economy, many of our children—as we know all too well—remain to be raised under conditions of severe sociocultural stagnation. Thus, we do not need to generate such "experimental" conditions in which the influence of sociocultural factors is kept to a minimum or is held constant. These conditions exist in our society anyway and—for better or worse—allow us to further explore the implications of Schaie's model. Hopefully, the fate of these children also makes us aware that our sciences, as much as our children, do not develop in a sociocultural vacuum; that it is naive, irrelevant, and irresponsible to anchor our efforts upon an abstract truth criterion traditionally conceived as either "god-given" or as provided by the "scientific facts" and "nature itself." Truth and knowledge are also functions of the actions taken and of the actions that they demand. These actions are determined by the economic and political ideologies of the societies in which we live. It has been the major purpose of this paper to make us cognizant of these dependencies so that we might better appreciate the alternative viewpoints of developmental science and, perhaps, synthesize them at higher theoretical levels, that is, levels at which not only different interpretations of individual development are considered but those of cultural-historical development as well.

## REFERENCES

Aebli, H. *Didactique psychologique: Application à la didactique de la psychologique de Jean Piaget.* Neuchâtel, Switzerland: Delachaux and Niestlé, 1951. (Republished: *Psychologische Didaktik.* Stuttgart: Klett, 1962).

Baltes, P. B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development,* 1968, 11, 145–171.

Barker, R. G., & Wright, H. F. *One boy's day.* New York: Harper, 1951.

Beard, M. *A history of business.* Ann Arbor: University of Michigan Press, 1962.

Berlyne, D. E. Recent developments in Piaget's work. *British Journal of Educational Psychology,* 1957, 27, 1–12.

Brown, R. *Words and things.* Glencoe, Ill.: Free Press, 1958.

Brunswik, E. *Systematic and representative design of psychological experiments.* Berkeley: University of California Press, 1949.

Chamberlain, H. S. *Die Grundlagen des 19. Jahrhunderts.* [The foundations of the 19th century] München, Germany: Bruckmann, 1909. (Translated by J. Lees, London: Lane, 1910).

Charles, D. C. Historical antecedents of life-span developmental psychology. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Theory and research.* New York: Academic Press, 1970.

Chomsky, N. *Syntactic structures.* The Hague: Mouton, 1957.

Chomsky, N. *Aspects of the theory of syntax.* Cambridge, Mass.: M. I. T. Press, 1965.

Darwin, C. *On the origin of species by means of natural selection, or, the preservation of favored races in the struggle for life.* London: Murray, 1859.

Darwin, C. Biographical sketch of an infant. *Mind,* 1877, 2, 285–294.

Flavell, J. H. *The developmental psychology of Jean Piaget.* New York: van Nostrand, 1963.

Galton, F. *Hereditary genius: An inquiry into its laws and consequences.* London: Macmillan, 1869.

Galton, F. *Inquiries into human faculty and its development.* London: Macmillan, 1883.

Gobineau, J. A. *Essai sur l'inégalité des races humaines.* [The inequality of human races] (5th ed.) Paris: Firmin-Didot, 1884 (Translated by A. Collins, London: Heinemann, 1915).

Hall, G. S. *Adolescence.* New York: Appleton, 1904.

*Herodotus* (Translated by A. D. Godley) London: Heinemann, 1931.

Hobbes, T. *Elementa philosophica de cive.* Amsterdam: Elzevirium, 1669.

Holzkamp, K. Wissenschaftstheoretische Voraussetzungen kritisch-emanzipatorischer Psychologie: Teil I. *Zeitschrift für Sozialpsychologie,* 1970, 1, 5–21.

Hunt, J. McV. *Intelligence and experience.* New York: Ronald Press, 1961.

Hyatt, A. Genesis of the Arietidae. *Smithsonian Contributions to Knowledge,* 1890, 26 (673), 1–11.

Jensen, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review,* 1969, 39, 1–123.

Kantor, J. R. *Interbehavioral psychology.* Granville, Ohio: Principia Press, 1959.

Koga, Y., & Morant, G. M. On the degree of association between reaction times in the case of different senses. *Biometrika,* 1923, **15**, 355–459.

Lenin, V. L. *What is to be done.* New York: International Publishers, 1929.

Lewin, K. *Principles of topological psychology.* New York: McGraw-Hill, 1936.

McCullers, J. C. G. Stanley Hall's conception of mental development and some indications of its influence on developmental psychology. *American Psychologist,* 1969, **24**, 1109–1114.

Neary, J. A scientist's variations on a disturbing racial theme. *Life,* 1970, **68** (22), 64.

Nordenskiöld, E. *The history of biology: A survey.* New York: Knopf, 1928.

Payne, T. R. *S. L. Rubinstejn and the philosophical foundations of Soviet psychology.* New York: Humanities Press, 1968.

Pearson, K. On the laws of inheritance of man. *Biometrika,* 1904, **3**, 131–190.

Piaget, J. *The language and thought of the child.* New York: Harcourt and Brace, 1926.

Piaget, J. Explanation in psychology and psychophysiological parallelism. In P. Fraisse & J. Piaget (Eds.), *Experimental psychology: Its scope and method.* Vol. I. New York: Basic Books, 1968.

Piaget, J. *Structuralism.* New York: Basic Books, 1970.

Platt, J. What we must do. *Science,* 1969, **166**, 1115–1121.

Riegel, K. F. Development of language: Suggestions for a verbal fallout model. *Human Development,* 1966, **9**, 97–120.

Riegel, K. F. Some theoretical considerations of bilingual development. *Psychological Bulletin,* 1968, **70**, 647–670.

Riegel, K. F. History as a nomothetic science: Some generalizations from theories and research in developmental psychology. *Journal of Social Issues,* 1969, **25**, 99–127.

Riegel, K. F. Relational interpretation of the language acquisition process. In G. B. Flores d'Arcais & W. J. M. Levelt (Eds.), *Advances in psycholinguistics.* Amsterdam: North-Holland, 1970. (a)

Riegel, K. F. The language acquisition process: A reinterpretation of selected research findings. In L. R. Goulet & P. B. Baltes (Eds.), *Life-span developmental psychology: Theory and research.* New York: Academic Press, 1970. (b)

Riegel, K. F. Time and change in the development of the individual and society. In H. W. Reese & L. P. Lipsitt (Eds.) *Advances in child development and behavior.* Vol. 7. New York: Academic Press, 1972, in press.

Rousseau, J. J. *Si le rétablissement des sciences et des arts a contribué à épurer les moeurs.* Dijon, France: L' Académie Dijon, 1750. (Republished: *The miscellaneous works of Mr. J. J. Rousseau.* London: Becket and De Hondt, 1767).

Rousseau, J. J. *Emile ou de l'éducation.* [Emile or education] LeHaye: J. Néaulme, 1762. (Translated by B. Foxley, New York: Dutton, 1950).

Ruger, H. A., & Stoessiger, B. On the growth curves of certain characteristics in man (males). *Annals of Eugenics,* 1927, **2**, 76–111.

Ryder, N. B. The cohort as a concept in the study of social changes. *American Sociological Review,* 1965, **30**, 843–861.

Saussure, de F. *Cours de linguistique generale.* [Course in general linguistics] Paris: Payot, 1904. (Translated by W. Baskin, New York: McGraw-Hill, 1966).

Schaie, K. W. A general model for the study of developmental problems. *Psychological Bulletin,* 1965, **64**, 92–108.

Society for the Psychological Study of Social Issues. [SPSSI] Statement by SPSSI on current IQ controversy: Heredity versus environment. *American Psychologist,* 1969, **24**, 1039–1040.

Spencer, H. *The principles of psychology.* Vol. I. New York: Appleton, 1897.

Spranger, E. *Psychologie des Jugendalters.* Leipzig, Germany: Quelle und Meyer, 1924.

Stern, C., & Stern, W. *Die Kindersprache.* Leipzig, Germany: Barth, 1907.

Sumner, W. G. *Social Darwinism.* Englewood Cliffs, N. J.: Prentice-Hall, 1963.

Tuddenham, R. D. Soldier intelligence in World Wars I and II. *American Psychologist,* 1948, **3**, 149–159.

Uexküll, von, J. J. *Umwelt und Innenwelt der Thiere.* Berlin: Springer, 1909.

Vigotsky, L. S. *Thought and language.* Cambridge, Mass.: M. I. T. Press, 1962.

(Received November 16, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 2, 142–154

# ON THE IDENTIFICATION OF REPEATEDLY PRESENTED, BRIEF VISUAL STIMULI[1]

MICHAEL E. DOHERTY[2] AND STUART M. KEELEY

*Bowling Green State University*

Two hypotheses concerning the explanation of increasing perceptual performance with temporal repetition of tachistoscopically presented visual stimuli are discussed. R. N. Haber's conception in terms of increasing phenomenal clarity is criticized, primarily due to the confounding influence of the "stimulus error." An alternative explanation in terms of the perceptual independence of successive inputs is presented. The assumption of independence is built into a model that combines aspects of C. W. Eriksen's "clearest form" approach with Bayesian decision theory. Four items of support for independence are presented. Eriksen found (*a*) no effect of interstimulus interval on accuracy; (*b*) no difference in accuracy whether a form was preceded by the same or a different form, or by a flash. The present authors found (*c*) the lack of influence of the locus on an inconsistent stimulus in a sequence; and (*d*) quite accurate Bayesian predictions of multiple observation performance in three experiments using equal prior probabilities of the stimulus states.

Investigators in different laboratories have been examining the effects on perceptual performance of successively repeated, brief presentations of the same visual stimulus. The typical finding is of an increasing level of performance, measured by, for example, identification accuracy or rated clarity, as the number of opportunities ($n$) to observe increases. Responses may be to each individual observation or to sequences of observations. This article addresses two explanations of the effect of temporal repetition that are not only quite different but are also incompatible.

The *clarity hypothesis* which has been advanced vigorously by R. N. Haber and his colleagues, postulates that improvement in performance is mediated by increasingly clearer percepts over successive repetitions; that is, if one were able to measure "perceptual clarity" directly, the measure would be an increasing function of $n$.

The *independence hypothesis* holds that multiple-look performance increases because the subject is aggregating, or combining, successive observations that are perceptually independent; that is, if one were able to measure perceptual clarity directly, the functions relating the clarity measure to $n$ would have a zero slope.

In this article the word "independence" is used solely to refer to the assumption that the probability ($P$) of a particular perceptual response ($e$) conditional upon the $i$th stimulus presentation ($H_i$), is not influenced by perceptual responses to the previous stimuli in that sequence. Symbolically:

$$P(e_i/H_i) = P(e_i/H_i, e_1, e_2, \ldots, e_{i-1}). \quad [1]$$

## BOUNDARY CONDITIONS AND DEFINITIONS

Certain boundary conditions and definitions should be made explicit. Ruled out of consideration is research in which repetition is confounded with exposure duration, including classical psychophysical studies and the related microgenetic (Flavell & Draguns, 1957) research. Investigations involving repetitions of complex material (e.g., Wyatt & Campbell, 1951) are beyond the scope of this presentation, as is work dealing with different sense modalities (e.g., Swets & Birdsall, 1967). Investigations, including Pylyshyn (1965) and Smith and Carey (1966), which used multiple-look paradigms, but which were not interpreted in terms of either hypothesis and which do not bear directly upon either, are not treated.

---

[1] This investigation was supported by Public Health Service Research Grant MH-17487-01 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Michael E. Doherty, Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43403.

A *repetition*, or *observation*, is used to refer to a single tachistoscopic exposure of a visual stimulus. A *trial* is a series of *n* repetitions.

## Evidence for the Clarity Hypothesis

The clarity hypothesis derives primarily from investigations that have called upon subjects to make phenomenal reports, and it rests especially on an argument with three components: (*a*) the empirical repetition effect; (*b*) operations employed by the experimenter to have subjects make phenomenal reports rather than problem solve; and (*c*) analyses of subjects' verbal reports. Several ancillary arguments, which Haber conceptualizes as converging operations, are presented and discussed after this tripartite argument is analyzed.

### *Three-Part Argument*

#### *Repetition Effect*

The discussion is restricted for the moment to those studies performed as tests of the clarity hypothesis and interpreted as at least partly supporting that hypothesis. Virtually every condition of each of these studies obtained an increasing function for some perceptual index as *n* increased. The longest interruption intervals were in excess of 8 seconds, ruling out luminance summation.

Three studies were performed with seven-letter words as stimuli. Haber and Hershenson (1965) presented these words under seven different exposure duration conditions, and instructed subjects to report the letters they perceived. For all conditions there was an increasing percentage of words in which all letters were correctly identified as *n* increased from 1 to 25. The biggest increment was in the first few repetitions. In a similar study, Haber (1965) tried to rule out a response process, or guessing, interpretation of the repetition effect by giving subjects prior knowledge of the word to be presented, and then having subjects report the letters perceived. The probability of reporting all the letters of a word increased for both frequent and rare words at both low and high exposure durations. Similar increases were found by Hershenson and Haber (1965) for English words and Turkish words spelled in English. This block of studies is summarized in Haber (1967). Hershenson (1969) manipulated the stimulus structure of seven-letter arrays, using from zero-order to third-order approximations to English as well as English words. In Experiment I, 140 arrays served as stimuli, with the maximum *n* being 15. The subjects reported letters perceived on each repetition. Experiment II was comparable except that prior knowledge was introduced to control for "nonperceptual factors." Hershenson (1969) obtained a repetition function for both experiments, but found that with prior knowledge the repetition effect was "negligible after the second exposure [p. 330]." He concluded that "unintentional guessing" was going on, and expressed doubt as to whether increasing clarity had been demonstrated beyond the second look.

To rule out a possible alternative to the clarity explanation in terms of changes in the subjects' eye fixations, a series of experiments presenting single letters was performed. Haber and Hillman (1966) presented 1 letter at a time, out of an alphabet of 25 letters, for from one to five repetitions. The subjects reported letters, pieces of letters or flashes of light, and confidence judgments. Two dependent variables were plotted: percentage correct and a derived score that took into account confidence judgments, correct parts of a letter, and good guessing. Both indexes showed a generally increasing slope as a function of *n* from one to four, with a dip from fourth to fifth looks. Dainoff and Haber (1967) manipulated exposure duration and *n* with single-letter stimuli. Percentage correct identifications increased from one to nine repetitions. In Standing, Sales, and Haber (1968), luminance, as well as number of repetitions, was varied, such that the effects could be assessed separately. For the first time, only a forced-choice response was used. In previous studies subjects had been instructed to report a letter only if they had seen it, whereas in this study, subjects were instructed to guess which letter had been presented on each observation. Percentage correct increased over nine repetitions, with luminance influencing slope. This study was interpreted as ruling out an explanation of the repetition effect in terms of the subject's "willingness to hazard a guess as the repetitions of a given stimulus proceed."

Pomerantz, Kaplan, and Kaplan (1969) presented single-letter stimuli from an 18-letter alphabet for up to five flashes, investigating possible satiation effects. Experiment I subjects made identification responses and ratings of phenomenal clarity. In Experiment II, subjects had prior knowledge and made clarity ratings only. The investigators interpreted their results within the framework of a Hebbian neuropsychological model, but their explanation for increases in rated clarity, monotonic in Experiment I and increasing then in Experiment II, is clearly in the context of the clarity hypothesis. No data concerning the identification responses were reported.

Standing, Sell, Boss, and Haber (1970) used manipulations involving imagery and subvocalization of correct and incorrect single letters, as well as prior knowledge, to investigate the effect of "internal representations" upon recognitions. They had subjects rate letter clarity on a 5-point scale, across five repetitions. All conditions showed predominantly increasing functions, with the average increase being about 10% of the total scale range. The increases were interpreted as due to increasing clarity mediated by progressively greater arousal of the appropriate cell assembly (see Haber, 1969). Haber, Standing, and Boss (1970) found that with a forced-choice response, the probability of a correct response increased with *n* even though the typeface and retinal locus of the form varied. For the first time, the experimenters explicitly rejected the position that increasing clarity is mediated by "simple sensory facilitation," and again discussed the effect in terms of cell-assembly theory.

Two investigations were not interpreted unambiguously in favor of the clarity hpothesis.

Standing and DaPolito (1968) employed an indicator methodology (Sperling, 1960) with arrays of letters. The primary data were the number of letters reported in indicated rows. They found a repetition effect when subjects withheld their response until after the final observation, but no increase when subjects responded after each observation, results that they interpreted in terms of response-produced disruption of short-term memory.

Earhard and Fullerton (1969) reported on three studies dealing with the clarity hypothesis. Their Experiment I, for example, called for subjects to identify letters in four-consonant arrays in one condition, to make "clearly seen" responses in another, and, in a unique condition, to state only the locus of a letter drawn with thin lines. Plotting proportion of letters correctly identified over 24 repetitions, they obtained a characteristic clarity function. Plotting proportion of letters "clearly seen" against *n*, they obtained a slope of .0055, which differed significantly from zero. But the function for the location of the thin letter was not significantly different from zero, being slightly negative for all four subjects. Earhard and Fullerton (1969) were apparently somewhat ambivalent in interpreting their findings. In discussing a study by Harcum, they seemed to reject Haber's position when they said:

> these data confirm those of the present study and offer little support for the position that repeated *successive* presentations improve the clarity of a stimulus term [p. 107].

But they also concluded that "repetition can increase the degree to which visual details of the physical stimulus can be discriminated only to a slight degree [p. 108]." The question of theoretical significance, however, is not whether the "degree" is slight, but whether repetition per se enhances a percept.

### *Operations Influencing the Subject's Report Behavior*

The second component that is indispensable to the clarity interpretation of verbal report data is some operation, typically instructional, that putatively provides that the subjects are reporting what they see, not what they infer is there. This operation is presented in detail for those experiments presented immediately above as firm evidence for the clarity hypothesis.

Five studies using complex stimulus displays are discussed first, concentrating exclusively on those aspects relating to the elicitation of phenomenal reports. The relevant instructions in Haber (1965), Haber and Hershenson (1965), Hershenson and Haber (1965), and Hershenson (1969), Experiment I, all of which used seven-letter words, are described similarly in the four reports. The study by Haber and Hershenson (1965) is somewhat more complete:

> For each stimulus exposure, *S* [the subject] was required to report verbally the letters that he was certain he perceived and their positions, even when he was certain of the word. This was given great stress in the instruction. All of the analyses here are based on the identification of letters, not of words. The *S* was said to have perceived the word if he correctly identified all of the seven letters on any of the exposures. This criterion, using letters rather than words, maximizes the probability of reports based on what *S* saw, rather than on what he thought he saw, or what he was sure must have been presented, but did not actually see [p. 41].

Four experiments used single-letter stimuli. Haber and Hillman (1966) described the relevant instructions as follows:

> The *E* [experimenter] emphasized that the experiment was primarily concerned with what the *S* saw, not what he guessed he saw. The *S* was therefore urged to report as best he could whether a letter, part of a letter, or just a flash of light was perceived. *He was also asked to indicate his confidence that what he saw was in fact what was there* . . . [italics added; p. 348].

Dainoff and Taber (1967) stated: "The *S*s were explicitly instructed not to guess, but to report exactly what they saw. A problem solving set was discouraged [p. 133]."

Pomerantz et al. (1969) had each subject identify the stimulus and also instructed them as follows:

> judge the clarity of her percept of the stimulus using three categories, "nothing" if no part of the stimulus was detected, "vague" if part but not all of the stimulus was perceived, and "clear" if the entire stimulus was perceived [p. 130].

Standing et al. (1970) employed a rated clarity dependent variable, emphasizing that

> the clarity ratings should be given solely on the basis of the sharpness or blackness of the perceived form and not on the *S*'s confidence of recognition of the form [p. 89].

### *Subjects Verbal Reports*

Haber (1965) and Haber and Hershenson (1965) stated that the analysis of the subject's recognition performance across repetitions, and also the subjects' own verbal reports, left no doubt that there was "a slow development of a percept." In other words, evidence for clarity is adduced by Haber and his colleagues from the subject's own introspective comments as well as from the evaluation of the experimental results described.

These three factors, taken in concert, form the essential evidence for the clarity hypothesis, and are evaluated before ancillary evidence is discussed.

## *Evaluation of the Three-Part Argument*

### *Repetition Effect*

There is absolutely no question that identification performance increases over repetitions, given the conditions described above. The question is whether alternatives to the clarity explanation have been ruled out by adequate controls.

### *Operations Influencing the Subjects' Report Behavior*

With respect to the operations used to elicit phenomenal reports, the present authors disagree with Natsoulas (1967) and contend that the operations were not adequate. Placing the problem into its historical context, Haber and his colleagues have essentially *instructed* subjects to avoid the *stimulus error,* which, as was demonstrated by the structuralists, is extraordinarily difficult for the subject to do. The instructions used in the clarity research could have had but little effect on the subjects' lifelong habit of adopting a problem-solving attitude toward the environment; on the subjects' habit, so powerfully reinforced by the language community, of trying to report what is *out there,* rather than on the contents of one's immediate experience.

Furthermore, some aspects of the methodology employed probably enhanced the tendency of subjects to respond by problem solving. While letters and words are relatively simple stimuli, they are in fact multifeatured, highly overlearned stimuli, with which subjects had had extensive experience of a problem-solving sort. Even the Hershenson and Haber (1965) "Turkish words" do not escape this criticism, since the "unfamiliar" terms were printed with English graphemes, and were pronounceable in English.

Another procedure that may have enhanced a problem-solving set was permitting subjects to identify the stimulus. Not only did they permit subjects to report what the stimulus was, they asked each subject for his confidence that what he saw was in fact there (cf. the Haber & Hillman, 1966, instructions above) and sometimes told him what the stimulus had been (e.g., Dainoff & Haber, 1967; Haber & Hillman, 1966). The fact that subjects would often misname the stimulus provides further evidence that the subjects were concerned with what the *stimulus* was, rather than with describing their percepts. The criticism is not met by analyzing individual-letter data, since the concern is not with what the experimenter does, but with what the subject is likely trying to do. Similarly, having the subject verbally name the distal stimulus or even features of the distal stimulus may also have enhanced the tendency to problem solve. The problem is not avoided by employing "rated clarity" as a dependent variable, since the subject is becoming progressively more certain of the identity of the stimulus, and his reports might reflect increasing certainty rather than increasing clarity. Instructions simply do not do the job. The interpretation of these verbal reports is ambiguous in the absence of a network of converging operations.

At this point the present authors would like to state as emphatically as possible that "making the stimulus error" and "guessing" are very different. Haber has often rejected "guessing" as a source of variance (e.g., Haber, 1969). But to ignore the possibility that the subjects are making the stimulus error is another matter entirely. Making a guess presumably involves some awareness on the subject's part that he does not know an answer and that he is responding in relative ignorance. The "stimulus error" is of the very warp and woof of cognitive activity. It has been a problem for perceptual psychologists since the beginning of scientific psychology. In fact, in the first issue of the first psychological journal in the English language, the issue was already joined. Lewes (1876) wrote: "For any writer to speak of sensation when he means sensation *plus* judgment would be as reprehensible as to speak of oxygen when he meant carbonic acid [p. 161]." The present authors believe that the operations herein criticized, including "clearly seen" responses and "rated clarity," are clearly "sensation *plus* judgment."

### *Subject's Reports*

The third aspect of the argument is the introspective report of the subjects. The present authors have run subjects and served as subjects in repetition experiments. Our own phenomenal experience with Landolt rings and graphemes ran counter to that which one would expect from a clarity hypothesis. Furthermore, Haber and Hillman (1966), attempting to show that guessing was not a viable explanation of their results, provided evidence from subjects' reports against the clarity hypothesis. They stated that "many reports of not seeing anything or of only pieces of the letter occurred after a trial in which *S* reported that he saw the entire letter clearly [p. 349]."

Introspective evidence is extraordinarily difficult to evaluate, and certainly cannot be taken as hard evidence for or against any specific scientific hypothesis. It is simply rejected as an argument for the "clarity hypothesis," or any other explanation of the repetition effect.

### *Ancillary Arguments for the "Clarity Hypothesis"*

Discussing the three earliest studies cited above that used wordlike stimuli, Haber (1967) asserted that "the effects of repetition are invariant over a number of quite different operations—duration, language, prior exposure, and frequency . . . [p. 209]." He buttressed his contention that the effect of repetition was directly upon perceptual clarity by "extensive curve fitting operations performed on the repetition functions . . . [p. 209]." First, let us address the curve-fitting operations.

### *Goodness of Fit*

The invariance referred to is reflected in the regularity of the function from one study to another, and in the fact that the same equation fits the functions from several of the earlier experiments. This equation is:

$$Pn = A - Q_1/n, \qquad [2]$$

where $Pn$ is the probability of a correct response on the $n$th observation, $A$ is the asymptote, and $Q_1$ is the probability of being incorrect on the first observation. It is difficult to evaluate the functions from the word studies for several reasons. First, the plots are of the probability of perceiving a word, while the dependent variable is number of letters reported. Second, the function has two parameters, $A$ and $Q_1$, free for fitting. The final and really salient point, however, is that the curve-fitting operations are irrelevant to the clarity hypothesis, and simply reflect the fact that the repetition effect is regular and somewhat transsituational. This calls attention to the essentially nonquantitative character of the clarity hypothesis, which precludes any theoretically grounded quantitative prediction of multiple-observation performance from single-observation performance until associated retinal or neural changes are specified quantitatively. Even if one were to grant the assertion that "the entire function can be generated by giving only one trial per word . . . [Haber, 1967, p. 210]," such prediction would be based upon purely empirical grounds.

### *Duration, Language, Frequency, Prior Knowledge*

The fact that the repetition effect obtains with different exposure durations, with foreign words spelled in English, and with English words of different Lorge-Thorndike counts is not relevant to the possibility that successive observations are independent. Nor is prior knowledge an adequate control. Haber (1965) argued that a response process interpretation of the repetition effect should "predict that repeated exposures should have no effect on recognition [p. 283]" when subjects have prior knowledge. This is not a necessary consequence of a response process interpretation, since it ignores what the subject is trying to do. While it has been argued that it is virtually impossible to perform the task demanded by Haber et al. (1970), the subjects were very likely trying to follow the experimental instructions to the best of their ability. Their attempts to give phenomenal reports would presumptively be based upon their phenomenal experience and their inferences from that experience. Again, it is argued that it is simply not that easy to obviate the stimulus error.

In summary, the major arguments for the clarity interpretation of the repetition effect have all been seriously questioned.

## EVIDENCE FOR THE INDEPENDENCE HYPOTHESIS

Other investigators have asked subjects to identify what is "out there," and have assumed that the subject's response represents the combining of independent perceptual responses from all observations on a trial. These authors also find an increase in identification accuracy as a function of $n$.

Before proceeding to the studies, the major similarities and differences between Haber's experimental procedures and those of the studies cited are noted. The same basic paradigm, employing temporal repetition with briefly presented stimuli in the visual modality with highly practiced subjects, is used. There are several differences, but none precludes inferences from a given position from being relevant to both. For example, the studies favoring the independence hypothesis have employed small alphabets, three or four stimuli, while Haber et al. have employed large alphabets, as noted. The most important methodological difference lies in the dependent variable. In Haber's studies, subjects respond after each observation. In the independence studies the method is what Green and Swets (1966) described as a fixed-observation procedure. The subjects are given a predetermined number of observations, their responses are made after the final observation, and they are instructed to identify what is "out there."

*Eriksen's Research*

The first study of direct relevance is Eriksen and Hoffman (1963). They presented single letters selected from the alphabet A, T, U for either one or two repetitions. They varied interstimulus interval from 0 to 450 milliseconds, as well as several other parameters, in addition to number of repetitions. Forced-choice identification accuracy to two repetitions was systematically superior to that for one repetition. The question of whether there had been some perceptual carry-over from observation to observation in the two-repetition trials, or whether the successive observations had been independent, was dealt with by post hoc analysis of the data. Eriksen and Hoffman (1963) raised, but did not purport to answer the question of independence, subsequently attacked more directly by Eriksen.

Eriksen (1966) performed three experiments that he interpreted as establishing the independence of successive inputs. The first two were similar to Eriksen and Hoffman (1963), but they extended the interstimulus interval range to include from 0 to 1,500 milliseconds. Experiments I and II both involved one and two repetitions of the same stimulus, while Experiment III presented two stimuli that were as likely to be different from one another as they were to be the same. From Experiment I, Eriksen argued for independence by comparing the empirical results for two repetitions with predictions of these results from a model that assumed independence.

Paraphrasing Eriksen (1966, pp. 30–31), the model assumed that the presentation of a stimulus resulted in one of three perceptual states, A, B, or C, corresponding to confidence ratings of "very confident," "think so," and "guess," respectively. The one observation condition allows the assessment of the probability of occurrence of each state, given each stimulus, and also the proportion of correct responses, or hit rate associated with each state. These one-look probabilities were multiplied together to calculate the probabilities of all possible multiple-look combinations of perceptual states, thus explicitly assuming independence. All combinations containing at least one A state were weighted by the one-observation A-state hit rate, and summed. All combinations containing a B state but no A state were weighted by the one-observation B-state hit rate, and summed. Finally, the one combination containing only C states was weighted by the C-state hit rate. These weighted sums were then summed to obtain a predicted hit rate, which was compared with the empirical hit rate. Note that multiplying the likelihood of a combination by the hit rate associated with the clearest perceptual state in that combination involves an assumption about a decision rule, that is, that the subject's judgment is determined by the single perceptual state corresponding to the most confident implicit response represented in the combination. Thus, on the decision side, Eriksen's model is a "clearest-state" or "best-look" model of the combination of observations. Using group data, the predicted mean hit rate for the Experiment I data, for interstimulus intervals up to 100 milliseconds, was .618, and the obtained was .611.

Experiment II was similar to Experiment I, except that interstimulus interval was extended to longer intervals. The evidence for independence involves the function relating percentage of correct identifications to interstimulus interval of from 50 to 1,500 milliseconds. Referring to that function (his Figure 3), which shows neither a systematic increase nor decrease, Eriksen (1966) stated:

> Lack of evidence for an increasing advantage of the double stimulation over lag intervals as long as 1500 msec. supports the conclusion of Exp. 1 that successive form stimulations separated by an interval of less than 1 msec. are independent [p. 33].

Eriksen's (1966) Experiment III provides yet another test of perceptual independence. Using short intervals and presenting only the forms A and U, Eriksen presented all four possible combinations of the two forms as well as one-look trials. The subjects stated what each stimulus had been, or if only one had been presented what it had been and which interval it had come in. Identification accuracy was not influenced by whether a form was preceded or followed by a form or a flash, suggesting that there was no interaction between the successive stimulations.

Eriksen concluded that these three studies were strong evidence for independence.

In light of Eriksen's data, the present authors used a successive observation condition as a performance parity comparison in two parallel processing studies (Keeley & Doherty, 1968, 1969). The first of these had two experiments, each with three conditions. Experiment I used Landolt rings as stimuli, and Experiment II used the letters A, T, and U. The three conditions were one observation, four successive observations, and four simultaneous observations. The relevant finding is the consistent underprediction of the successive observation data by Eriksen's clearest state model. This led the authors to criticize the decision aspect of this model, on the grounds that the upper bound of the predicted multiple-observation hit rate was the single-observation A-state hit rate. The Eriksen model predicts multiple-observation performance from a single, inferred "best look." Borrowing from Eriksen's thinking, and from a decision theory, the authors (Doherty & Keeley, 1969) proposed a simple application of Bayes' rule, which assumes that the subject is aggregating all of the observations he makes, and which predicts his total response pattern rather than some overall index.

The Bayesian model requires the estimation of the probability of all possible combinations of perceptual responses, contingent upon the individual stimuli. Bayes' rule is used to calculate the most likely state of the world for each possible combination of perceptual responses. This most likely state of the world is predicted to be the subject's response to a given combination. Appropriate summations of conditional probability values yield a predicted stimulus–response matrix. The decisions based on optimally aggregated observations are then compared with the subject's multiple-look identification responses. Note that the probability of the joint occurrence of the responses within a set of multiple-observation implicit responses is inferred by multiplying together the one-look conditional probabilities of the individual responses, given the stimuli; and that the aggregation is performed according to the form of Bayes' theorem appropriate to independent events.

### *Two Investigations of a Bayesian Model*

In the first study (Doherty & Keeley, 1969), two subjects each served 40 hours identifying Landolt rings differing only in gap orientation. The subjects performed in one- and four-observation conditions. No parameters were free for fitting in the prediction of each subject's pattern of multiple-observation performance. The predicted and obtained matrices were very similar. The overall predicted hit rates for these subjects were very close to the empirical hit rates, differing by only .004 and .039. Experiments in which values of the behavioral measure are predicted involve acceptance of the null hypothesis. However, such prediction, especially without free parameters, is certainly one of the more stringent tests of a scientific model. The authors regard this study as rather strong support for the prediction model employed, and for the assumption of perceptual independence.

Keeley, Doherty, and Bachman (1970) employed three subjects for an extended time, presenting them with the graphemes A, T, and U in one- and four-observation conditions. The subjects made decisions about what stimulus state had led to the observation or observations they had experienced, but they made confidence ratings as well. Three sets of predictions were made for each subject, with different degrees of precision of the confidence ratings. All three sets of predictions, each set made without free parameters, matched the empirical probabilities well. The closest predictions for all three subjects were based on the prediction matrix closest in size to the 4 × 4 prediction matrix of the first study. The errors in hit rates predicted from the 3 Stimuli × 6 Response Matrices were .003, .003, and .018 for the three subjects. The average underprediction from the 3 × 3 matrix (without confidence ratings) was .034, the average overprediction using all four levels of confidence rating was but .023.

### *Two Conditions Degrading Decision Performance*

The accuracy of the prediction led the authors to consider the case for independence and for Bayesian aggregation fairly well established for the conditions common to these two highly similar experiments. The formal and mathematically explicit nature of the model led to a search of formal characteristics of the task yielding noncorrespondence between the subject's performance and mathematically optimal predictions. An obvious possibility was unequal a priori probabilities.

In Keeley and Doherty (1971a) seven subjects were presented Landolt rings varying in four orientations. For four subjects the gaps had probabilities of occurrence of .7, .1, .1, and .1; and for three subjects, .3, .3, .3, and .1. This contrasts with equal a priori probability values of .25 in Doherty and Keeley (1969) and .33 in Keeley et al. (1970). The predictions under the Bayesian model were not nearly so accurate as they had been in previous studies. For all four subjects in the .7, .1, .1, .1 condition, the model systematically overpredicted the hit rate to the .1 stimuli, and underpredicted the hit rates to the .7 stimuli. The results for the other condition were less systematic, but the predictions for the .3 stimuli were reasonably accurate.

The Bayesian model did not predict well in this study, but the independent variable was explicitly selected in the anticipation that it might point to an environmental condition that renders decision performance nonoptimal.

The results however are incompatible with the clarity hypothesis. The repetition effect was markedly different for the stimuli that occurred .7 of the time as compared with those that occurred .1 of the time. There was a substantial increase in identification accuracy for the .1 stimuli, the mean hit rate going from .18 with one observation to .44 with two observations. On the other hand, the mean hit rate for the .7 stimulus was .81 for one, two, and four observations, showing no change as a function of $n$. The clarity interpretation predicts not only the observed large change with the .1 stimuli but also an increasing function for the .7 stimuli, since the starting identification accuracy of .81 left ample room for improvement. There is no possible way to construe that hypothesis which would lead to an increase for some stimuli over repetitions but to no increase for other stimuli at the same exposure duration. There are conditions that, however, lead to predictions of no change, or even decrements, under a decision model.

Doherty and Keeley[3] sought to determine whether certain structural characteristics of the stimuli would lead to nonoptimal aggregation. Stimuli varying only unidimensionally (Miller, 1956) were decided upon, and an alphabet of straight lines varying in length was selected. Performance was degraded by using lines that were not perfectly identifiable under good viewing conditions and by manipulating the exposure duration to obtain a desired one-look hit rate. Thus, the stimuli were both state and process limited (Garner, 1970). There were radical differences in the frequencies of response, presumably because of the subject's tendency to give the response corresponding to the shortest stimulus when only a piece of the horizontal line was perceived. The statistical characteristics of the prediction matrices resulting from this bias made this investigation analogous to the unequal a priori study cited. Overall hit rates were well predicted, but there were sizable discrepancies in the predictions of the individual responses. Again there was a differential change over repetitions for different stimuli within the alphabet. The short lines showed no improvement in identification accuracy over repetitions. Conversely, the long lines showed the increase generally expected with repetitions. These changes are compatible with those to the .7 and .1 stimuli of the previous study.

There are theoretical difficulties involved when one tries to establish independence by matching the subject's behavior with a normative model, especially when it is assumed that optimal performance will obtain only under certain sets of conditions. A great deal of empirical research would be needed to specify both those conditions leading to optimal performance and those degrading performance. A great deal of theoretical work would then be required to incorporate those conditions into the same conceptual system, such that the model would either hold that observations under all conditions are independent and aggregation is nonoptimal or that specified conditions lead to a specified form of dependency. Furthermore, as Garner and Morton (1969) have pointed out, correlated input procedures, of which the repetition paradigm is one, typically provide only indirect evidence for or against the assumption of independence. Using such a procedure, assumptions must be made not only about the subject's perceptual processes but also about his decision processes. Predictions may differ depending upon the decision rule assumed, and therefore good prediction in any given study is not compelling evidence for independence. Thus, two studies were performed dealing directly with the two hypotheses in question. One provides further evidence for the Bayesian model, as well.

---

[3] M. E. Doherty & S. M. Keeley. Identification accuracy to repeated presentations of pure tones and straight lines. Paper presented at the meeting of the Midwestern Psychological Association, Cincinnati, May 1970.

### *Two Investigations Aimed at Testing the Hypotheses*

The first of these studies (Doherty, Keeley & Norton, 1971) tested a deduction from the clarity hypothesis. This study had operations, on both the independent variable and dependent variable sides, aimed at attenuating the subject's tendency to make the stimulus error. Landolt rings were used because they are relatively nonmeaningful and single featured, thus not susceptible to "filling-in" by subjects on the basis of redundant information. An additional operation on the stimulus side involved physically degrading 50% of the stimuli by randomly masking the stimuli on half of the trials such that the subject could see only part of a Landolt ring. Thus the subject should not have been surprised when he saw only part of a ring. It was argued that the use of the language system might enhance the subject's tendency to problem solve; thus subjects who viewed the Landolt rings drew pictures of what they saw, rather than respond verbally. A second group of subjects then viewed their reproductions and made decisions concerning that which the first subjects had been presented.

The decision-making subjects looked at different numbers of reproductions, such that decisions made on the basis of all the drawings of a subject on each trial could be compared with an equal number of first-observation drawings, second-observation drawings, etc., from those trials. Thus, the number of reproductions was held constant as the order of the observation (first, second, third, and fifth) varied. Given that a drawing is as veridical a representation of the state of the subject's sensorium as a verbal response, then the clarity hypothesis predicts that decisions based on drawings from observations later in a trial would be more accurate than decisions based on drawings from observations earlier in the trial. The fact that subjects had to look away and draw a response should not make any great difference, otherwise the clarity interpretation would be limited to the situation in which a subject is tachistoscopically presented repeated opportunities to observe, but only so long as he is looking into that tachistoscope. Twenty-one Stage II subjects made decisions based on drawings from 3 Stage I subjects. The data from 2 of the 3 Stage I subjects ran counter to the clarity predictions, with no systematic increases in accuracy as a function of the order of the observation. The third subject seemed to be unable to avoid the stimulus error. His drawings showed a remarkable perseveration; he drew successive, complete Landolt rings even though the ring being drawn did not correspond to the one he was being presented.

This study was interpreted as being a flat rejection of the clarity hypothesis, but the rather elaborate technques aimed at attenuating the stimulus error permitted the collection of only a very small amount of data, even though the Stage I subjects served for 10 hours each after pretraining of several hours.

A complementary study (Keeley & Doherty, 1971b) provided a further test of the Bayesian model with equal prior probabilities, and also a rather direct test of the hypothesis of perceptual independence. If successive observations are independent, then the order in which these observations appear should be irrelevant to the final decision. In the study under consideration, many sequences of stimuli had objectively inconsistent information. The stimulus forms A, T, and U were used, and the subject would have one of these presented on a one-look trial. But the multiple-observation trials consisted not only of the sequences AAA, TTT, and UUU but also sequences such as AAT, ATA, and TAA, but the subject was required to make a single-letter response. Three-observation trials corresponding to all possible combinations of A, T, and U, except the sequence ATU, were presented to subjects. The decisions made by subjects speak to the question of independence by providing answers to two questions. First, is there any consistent effect due to the locus of the objectively inconsistent stimulus in the sequence? Second, does the formal model assuming independence, that is, Bayes' rule, predict the subject's multiple-observation performance well? Collapsing across sessions, subjects, and specific inconsistent letter, the hit rates for sequences with inconsistent letter first, second, and third were .55, .54, and .54, respectively. The *F* ratio for order was 1.26 ($df = 2$ and 10). The hit rate for two subjects decreased slightly as the inconsistent letter shifted from first to third position in the sequence, one increased then decreased, the other three showed virtually identical hit rates for the three loci.

TABLE 1

EMPIRICAL (EMP) AND PREDICTED (PRE) PERCENTAGES OF EACH RESPONSE TO THE NINE STIMULUS COMBINATIONS, AVERAGED OVER SUBJECTS, AND AVERAGED OVER ORDERS OF EACH COMBINATION

| Responses | Emp | Pre | Emp | Pre | Emp | Pre |
|---|---|---|---|---|---|---|
| | AAA | | TTT | | UUU | |
| a | 72 | 72 | 16 | 15 | 14 | 17 |
| t | 14 | 13 | 71 | 70 | 13 | 13 |
| u | 14 | 14 | 12 | 15 | 73 | 70 |
| | AAT | | TTU | | UUA | |
| a | 50 | 56 | 16 | 18 | 28 | 32 |
| t | 16 | 15 | 51 | 52 | 14 | 15 |
| u | 34 | 29 | 33 | 30 | 58 | 54 |
| | AAU | | TTA | | UUT | |
| a | 53 | 56 | 31 | 30 | 16 | 19 |
| t | 28 | 28 | 54 | 53 | 28 | 27 |
| u | 19 | 16 | 15 | 16 | 57 | 54 |

When the single-observation data of this study are used to predict multiple-observation performance, the same picture as in the earlier studies with equal a priori probabilities emerges. Not only are the overall hit rates predicted well, but the overall pattern of responding by each individual subject is predicted quite accurately. The number of comparisons in the multiple-observation condition precludes detailed presentation of the data, but Table 1 provides the response pattern to each stimulus combination, averaged over subjects. This study thus provides not only a very stringent test of, and support for, a model assuming perceptual independence but also rather direct support for the independence hypothesis.

### *Additional Evidence*

Ulehla, Halpern, and Cerf (1968) and Halpern and Ulehla (1970) also performed repetition experiments, and interpreted their results in a framework closely related to the model presented above, that is, signal detection theory, thus supporting the assumption of independence.

The research employing simultaneous presentations of graphemes provides the final evidence for the independence hypothesis. Eriksen and Lappin (1965) and Keeley and Doherty (1968, Experiment II), for example, have found growth functions for simultaneous presentations similar to those found with successive. The results of these parallel processing investigations are interpreted in terms of independent perceptual channels that provide independent opportunities to observe. Thus, results from two similar methods of providing multiple opportunities to observe, simultaneous and successive presentations, can both be understood in terms of perceptual independence, while there does not seem to be any obvious way in which the clarity hypothesis would predict growth functions for simultaneous presentations.

## CONCLUSION

Considering all of the evidence, the authors conclude that the clarity hypothesis cannot be considered a general hypothesis accounting for the visual repetition effect. The case for perceptual independence seems, for simple stimuli in situations comparable to the ones employed, to be a strong one. The hypothesis that subjects aggregate the perceptual evidence optimally, by a process that can be described by Bayes' rule, seems established for equally probable stimuli.

There are several additional generalizations of interest. First, the demonstration that successive percepts are independent permits the continued use of successive presentations as a parity condition in investigations of the parallel processing hypothesis, provided the investigator is willing to assume that the decision rule used with simultaneous input is the same as with successive. Second, independence assured, perceptual experiments such as those used by the present investigators are defensible procedures for investigating decision processes. However, the authors believe that experiments like the ones performed in our laboratories are inefficient vehicles for investigating decision processes.

In a final analysis of the two competing theoretical positions that this article addresses, the stance adopted by the authors seems to be more in agreement with the data, the model proposed is a heuristic one, and the model proposed seems to represent a more parsimonious approach to the explanation of the repetition effect. The clarity hypothesis "explains" only those experiments designed to test it—and these are all similar. It does not relate a diverse set of psychological phenomena. On the other hand, the Bayesian model relates the repetition effect to a variety of other psychological phenomena, most obviously those addressed by statistical decision theory. Given the emphasis in contemporary perception on formal, quantitative models of perceptual processes, the demonstration of independence and the application of the Bayesian model also relate the empirical repetition effect to general perceptual theorizing. Finally, conceptualizing the repetition effect in terms of decision making or problem solving is in the tradition of perceptual theories, from Helmholtz' "unconscious inference" to Gregory's (1970) "perception is a kind of problem-solving."

## REFERENCES

DAINOFF, M., & HABER, R. N. How much help do repeated presentations give to recognition processes? *Perception and Psychophysics*, 1967, **2**, 131–136.

DOHERTY, M. E., & KEELEY, S. M. A Bayesian prediction of four-look recognition performance from one-look data. *Perception and Psychophysics*, 1969, **5**, 362–364.

DOHERTY, M. E., KEELEY, S. M., & NORTON, G. K. A test of the clarity interpretation of the repetition effect. *Cognitive Psychology*, 1971, **2**, 386–399.

EARHARD, B., & FULLERTON, R. How much does repetition facilitate perception? *Journal of Experimental Psychology*, 1969, **81**, 101–108.

ERIKSEN, C. W. Independence of successive inputs and uncorrelated error in visual form perception. *Journal of Experimental Psychology*, 1966, **72**, 26–35.

ERIKSEN, C. W., & HOFFMAN, M. Form recognition at brief durations as a function of adapting field and interval between stimulations. *Journal of Experimental Psychology*, 1963, **66**, 485–499.

ERIKSEN, C. W., & LAPPIN, J. S. Internal perceptual system noise and redundancy in simultaneous inputs in form identification. *Psychonomic Science*, 1965, **2**, 351–352.

FLAVELL, J. H., & DRAGUNS, J. A. A microgenetic approach to perception and thought. *Psychological Bulletin*, 1957, **54**, 197–217.

GARNER, W. R. The stimulus in information processing. *American Psychologist*, 1970, **25**, 350–358.

GARNER, W. R., & MORTON, J. Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin*, 1969, **72**, 233–259.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

GREGORY, R. L. *The intelligent eye*. New York: McGraw-Hill, 1970.

HABER, R. N. Effect of prior knowledge of the stimulus on word-recognition processes. *Journal of Experimental Psychology*, 1965, **69**, 282–286.

HABER, R. N. Repetition as a determinant of perceptual recognition processes. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form*. Cambridge, Mass.: M.I.T. Press, 1967.

HABER, R. N. Repetition, visual persistence, visual noise, and information processing. In K. N. Leiboric (Ed.), *Information processing in the nervous system*. Columbus, Ohio: Springer, 1969.

HABER, R. N., & HERSHENSON, H. Effects of repeated brief exposures on the growth of a percept. *Journal of Experimental Psychology*, 1965, **69**, 40–46.

HABER, R. N., & HILLMAN, E. R. The effect of repetition on the perception of single letters. *Perception and Psychophysics*, 1966, **1**, 347–350.

HABER, R. N., STANDING, L., & BOSS, J. Effect of position and typeface variation on perceptual clarity. *Psychonomic Science*, 1970, **18**, 91–92.

HALPERN, J., & ULEHLA, Z. J. The effect of multiple responses and certainty estimates on the integration of visual information. *Perception and Psychophysics*, 1970, **7**, 129–132.

HERSHENSON, M. Stimulus structure, cognitive structure, and the perception of letter arrays. *Journal of Experimental Psychology,* 1969, **79**, 327–335.

HERSHENSON, M., & HABER, R. N. The role of meaning in the perception of briefly exposed words. *Canadian Journal of Psychology,* 1965, **19**, 42–46.

KEELEY, S. M., & DOHERTY, M. E. Simultaneous and successive presentations of single-featured and multi-featured visual forms: Implications for the parallel processing hypothesis. *Perception and Psychophysics,* 1968, **4**, 296–298.

KEELEY, S. M., & DOHERTY, M. E. A note on the parallel processing of single-featured forms. *Perceptions and Psychophysics,* 1969, **6**, 81–82.

KEELEY, S. M., & DOHERTY, M. E. A Bayesian prediction of multiple-look identification performance from one-look data: The effect of unequal prior probabilities. *Perception and Psychophysics,* 1971, **10**, 119–122. (a)

KEELEY, S. M., & DOHERTY, M. E. Bayesian aggregation of successive visual inputs. *Journal of Experimental Psychology,* 1971, **90**, 300–305. (b)

KEELEY, S. M., DOHERTY, M. E., & BACHMAN, S. P. A Bayesian prediction of four-look recognition performance from one-look data: II. *Perception and Psychophysics,* 1970, **7**, 218–220.

LEWES, G. H. What is sensation? *Mind,* 1876, **1**, 157, 161.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 1956, **63**, 81–97.

NATSOULAS, T. What are perceptual reports about? *Psychological Bulletin,* 1967, **67**, 249–272.

POMERANTZ, J. R., KAPLAN, S., & KAPLAN, R. Satiation effects in the perception of single letters. *Perception and Psychophysics,* 1969, **6**, 129–131.

PYLYSHYN, Z. W. Information available from two consecutive exposures of visual displays. *Canadian Journal of Psychology,* 1965, **19**, 133–144.

SMITH, F., & CAREY, P. Temporal factors in visual information processing. *Canadian Journal of Psychology,* 1966, **20**, 337–342.

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs,* 1960, **74** (11, Whole No. 498).

STANDING, L., & DAPOLITO, F. Limitations of the repetition effect revealed by partial report. *Psychonomic Science,* 1968, **13**, 297–298.

STANDING, L. G., SALES, B. D., & HABER, R. N. Repetition *versus* luminance as a determinant of recognition. *Canadian Journal of Psychology,* 1968, **22**, 442–448.

STANDING, L., SELL, C., BOSS, J., & HABER, R. N. Effect of visualization and subvocalization on perceptual clarity. *Psychonomic Science,* 1970, **18**, 89–90.

SWETS, J. A., & BIRDSALL, T. G. Deferred decision in human signal detection: A preliminary experiment. *Perception and Psychophysics,* 1967, **2**, 15–28.

ULEHLA, Z. J., HALPERN, J., & CERF, A. Integration of information in a visual discrimination task. *Perception and Psychophysics,* 1968, **4**, 1–4.

WYATT, D. F., & CAMPBELL, D. T. On the liability of stereotype or hypothesis. *Journal of Abnormal and Social Psychology,* 1951, **46**, 496–500.

(Received November 19, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 2, 155–158

# RAPID EYE MOVEMENTS AND VISUAL IMAGERY DURING SLEEP[1]

DAVID KOULACK[2]

*University of Manitoba*

A number of early findings supported the notion that there is a correspondence between eye movements and the visual imagery of dreams. Subsequent studies, however, yielded contradictory results. It is suggested that eye movements might occasionally be related to the visual imagery of dreams but that the notion of a constant isomorphic relationship between the two is untenable. It is also noted that rapid eye movement (REM) bursts might correspond in other ways to dream content, that is, as an integral and parallel part of the total picture of Stage 1 REM activation.

Although dreaming has held long-standing fascination for philosophers and scientists it was not until the discovery of periodic rapid eye movements (REMs) by Aserinsky and Kleitman (1953, 1955) that systematic research into the area burgeoned into a full-fledged enterprise. It was found that REMs occurred in conjunction with an electroencephalograph (EEG) pattern of random, fast, low-voltage waves (Stage 1 sleep) and that they occurred systematically throughout the night in approximately 90-minute cycles (Dement & Kleitman, 1957a, 1957b).

It should be noted that in 1892 Ladd suggested that when we dream our eyes move in much the same fashion as they do in the waking state, while when we are not dreaming our eyes remain quiescent. Investigators were then able to test this intriguing notion by awakening subjects while they were in the midst of REM periods and asking them what had been going through their minds. Using this procedure, a number of researchers (Dement, 1955; Foulkes, 1962; Goodenough, Lewis, Shapiro, Jaret, & Sleser, 1965; Goodenough, Shapiro, Holden, & Steinschriber, 1959; Hobson, Goldfrank, & Snyder, 1965; Rechtschaffen, Verdone, & Wheaton, 1963; Wolpert & Trosman, 1958) obtained 69%–88% dream recall from REM periods. The discrepancies in amount of recall from study to study can in large part be accounted for by differing criteria for designating a report as a dream (e.g., a brief image versus a complete coherent account), differing methods of awakening (abrupt versus gradual), and individual differences in ability to recall dreams.

At the same time a number of studies concerned with the question of whether or not REMs themselves were related to the visual imagery of dreams were launched. If a relationship between REMs and visual imagery could be demonstrated, it would serve to strengthen the notion that dreams (or at least visual dreams) occur during REM periods and not at other times and would enhance the possibility of using REM sleep as an "objective" indicant of dreaming. This would make such long-standing questions—How often do we dream? How long are our dreams? When in the night do we dream? Are there age and sex differences in amount of dreaming?—amenable to empirical investigation.

Investigations by Dement and Kleitman (1957), Dement and Wolpert (1958), and Roffwarg, Dement, Muzio, and Fisher (1962) seemed to lend strong support to the notion that REMs correspond to the visual imagery of dreams. All of these studies involved awakening subjects after distinctive bursts of eye movements (horizontal, vertical, etc.) and an examination of content just prior to the awakening. Roffwarg et al. (1962) went even further; they attempted to predict the nature of the eye movements on the basis of the dream content. When dream recall was rated as vivid by the subjects they found that the experimenter was able to predict the direction of the eye movements throughout the dream with approximately 80% accuracy. For example, Subject 7 reported:

---

[1] This research was aided by bibliographic information obtained from the University of California at Los Angeles Brain Information Service Sleep Bulletin and was supported in part by the Medical Research Council, Canada, Grant MA-3929.

[2] Requests for reprints should be sent to David Koulack, Department of Psychology, University of Manitoba, Winnipeg, Canada.

> Right near the end of the dream I was walking up the back stairs of an old house. I was holding a cat in my arms . . . I walked up as a dancer would, holding my head high, and I glanced up at every step I took. [There were] five or six [steps]. I reached the head of the stairs and I walked straight over to a group of people about to begin a circle dance . . . I looked straight ahead at the person across from me. Then I woke up [Roffwarg et al., 1962, p. 239].

In this instance the experimenter correctly predicted five vertical upward eye movements as the dreamer walked up the steps and a few horizontal eye movements when she reached the top. The authors attributed instances of deficiencies in prediction to the experimenter's inability to master the spatial orientation of the dreamer or the experimenter's misinterpretation of parts of the dreamer's reports. Other studies (Berger & Oswald, 1962; Kamiya [3]) which demonstrated a positive relationship between the activity of the dreamer in the dream and the extent of his eye movements also supported the relationship notion.

Work with blind people seemed to provide further support for the REM visual imagery relationship notion. It was reasoned that congenitally blind people should have no visual imagery in their dreams and no REMs. Berger, Olley, and Oswald (1962) found that three subjects who had lifelong blindness, and two subjects who had been blind for over 30 years, had lost their ability to picture things, and had no visual imagery in their dreams also had no REMs during Stage 1 sleep. In contrast, three subjects who had been blind from 3 to 15 years and still had visual imagery in their dreams also had REMs during their cyclically recurring Stage 1 sleep. Although the authors acknowledged that these findings supported the REM visual imagery relationship notion, they suggested that a more parsimonious notion accounting for both REMs and concomitant bursts of "saw-toothed" frontal EEG waves would be a sudden change of neurophysiological conditions without reference to dream content. Another study (Offenkrantz & Wolpert, 1963) also reported that a single congenitally blind subject had dreams devoid of both visual imagery and REMs.

Almost at the same time as the REM visual imagery relationship notion was being advanced, new findings suggested that REM sleep was not the sole repository of dreams. By awakening subjects in the midst of non-REM (NREM) sleep, a number of investigators (Foulkes, 1962, 1967; Goodenough et al., 1959; Rechtschaffen et al., 1963) were able to obtain dream recall, albeit the percentage of recall was somewhat less than that obtained from REM periods, ranging from 23% to 54%. (The variability in NREM recall might be a function of similar factors as those effecting amount of REM period recall.) Although it was noted that in general NREM content is less vivid and involves less elaborate imagery and emotionality (Rechtschaffen et al., 1963), there seems to be little question that visual dreams can and do occur during NREM sleep. These findings suggested that if there is a relationship between REMs and visual imagery in dreams it is not a constant one.

In addition, a number of recent studies have reported results that are in contradiction to those of the earlier studies and the REM visual imagery relationship notion. Amadeo and Gomez (1966) and Gross, Byrne, and Fisher (1965) were able to demonstrate the existence of REMs during Stage 1 sleep in the congenitally blind by use of mechanical [4] rather than electrical measures. Fishbein, Schaumburg, and Weitzman (1966) found that REMs occurred in kittens reared in darkness with no waking visual experiences and presumably no visual imagery during sleep. The findings that REMs occur in newborns (Delange, Castan, Cadilhac, & Passouant, 1962; Monad, Dreyfus-Brisac, Morel-Kahn, Pajot, & Plussart, 1963; Roffwarg, Muzio, & Dement, 1966; Weitzman, Fishbein, & Graziani, 1965) who have extremely crude patterned vision (Fantz, 1963; Haynes, White, & Held, 1965) and in decorticate cats (Jouvet &

---

[3] J. Kamiya. Behavioral and physiological concomitants of dreaming. (Unpublished progress report) Washington, D. C.: National Institute of Mental Health, 1962.

[4] The corneo-retinal potential in some blind people is lacking or diminished so that REMs are not apparent when the usual electrical recording techniques are used. In such situations REMs can be recorded by tiny strain gauges.

Jouvet, 1963), the cerebral cortex being necessary for complex pattern vision, also tend to cast doubt on the notion of a constant isomorphic relationship between REMs and the visual imagery of dreams.

Perhaps the most weighty evidence against the relationship notion is found in the work of Jacobs, Feldman, and Bender.[5] They studied eight subjects over a total of 20 nights of sleep and monitored their eye movements by DC[6] electro-oculograms (EOGs), as well as visual inspection of the position of the eyes under closed lids. Subjects were awakened from 41 of the 71 REM periods recorded. Awakenings followed either bursts of eye movements or ocular quiescence, and dream narratives following the awakenings were graded according to the clarity with which the subject reported the last few events of the dream action. They found that only occasionally did the eye movements match the action of the dreams. Generally, the movements of the eyes did not match dream narratives, but rather, in spite of a great deal of variability in the content, there was a certain amount of consistency in the percentage of types of eye movements across all REM periods. The authors also noted that unlike waking eye movements, those occurring during REM sleep were torsional or oblique.

The question then is where does the REM visual imagery relationship notion stand, given this apparent contradictory data. Antrobus, Dement, and Fisher (1964) found that nonrecallers (people who have difficulty in remembering their dreams) have significantly more REMs (20%) than recallers (13%). Noting that Luborsky, Blinder, and Mackworth (1963) found that subjects looked away from threatening aspects of pictures and Antrobus, Antrobus, and Singer (1964) found that subjects tried to "look away" from imagined events when they were asked to suppress them, the authors suggested that nonrecallers might have more eye movements than recallers because they are looking away from frightening or threatening aspects of their dreams. While this interpretation might explain some of the lack of correspondence between REMs and visual imagery found by Jacobs et al. (see Footnote 5), that is, some subjects might at times try to "look away" from the imagery of their dreams, it is unlikely that it accounts for the entire discrepancy.

In a study of respiration patterns, Aserinsky (1965) advanced the notion that a neurophysiological mechanism might be responsible for both REM bursts and associated respiration rate increase. He noted that there is a periodicity in respiration rate change linked with REM bursts and that the REM bursts themselves are accompanied by increased occipital alpha rhythm, an occurrence that is counter to the reduction of alpha rhythm usually found as a concomitant to visual imagery in the waking state. The findings of Spreng, Johnson, and Lubin (1968) that REM-burst rate is consistent within subjects from epoch to epoch and night to night, that there is a consistent individual subject bias in eye movement direction, and that there are correlations between REM-burst time and autonomic activity also tend to point to a neurophysiological or neurochemical mechanism underlying the REMs themselves.

In sum then, the initial hypothesis of a constant isomorphic relationship between REMs and visual imagery seems to be untenable, although it is quite possible that from time to time eye movements and the visual imagery of dreams might be related. This does not mean, however, that REM bursts do not correspond in some other ways to dream content. The findings that REM density is positively related to such dream attributes as bizarreness (Goodenough et al., 1965; Verdone, 1965), vividness (Verdone, 1965), emotionality (Karacan, Goodenough, Shapiro, & Starker, 1966; Verdone, 1965), aggression (Karacan et al., 1966), and intensity (Pivik & Foulkes, 1966) suggest the possibility that Stage 1 dreams and the eye movements that accompany them are integral and parallel parts of the total picture of activation during Stage 1 REM sleep.

---

[5] L. D. Jacobs, M. Feldman, & M. B. Bender. The pattern of human eye movements during sleep. Paper presented at meeting of the American Neurologic Association, June 1970.

[6] Generally EOGs are AC recordings which only permit monitoring of the movements. The DC EOGs permit monitoring of both movements and position of the eyes and hence are ideal for examination of the REM visual imagery relationship notion.

## REFERENCES

Amadeo, M., & Gomez, E. Eye movements, attention, and dreaming in subjects with lifelong blindness. *Canadian Psychiatric Association Journal*, 1966, **11**, 500–507.

Antrobus, J. S., Antrobus, J. S., & Singer, J. L. Eye movements accompanying daydreaming, visual imagery, and thought suppression. *Journal of Abnormal and Social Psychology*, 1964, **69**, 244–252.

Antrobus, J. S., Dement, W., & Fisher, C. Patterns of dreaming and dream recall. *Journal of Abnormal and Social Psychology*, 1964, **69**, 341–344.

Aserinsky, E. Periodic respiratory pattern occurring in conjunction with eye movements during sleep. *Science*, 1965, **150**, 763–766.

Aserinsky, E., & Kleitman, N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 1953, **118**, 273–274.

Aserinsky, E., & Kleitman, N. Two types of ocular motility occurring in sleep. *Journal of Applied Physiology*, 1955, **8**, 1–10.

Berger, R. J., Olley, P., & Oswald, I. The EEG, eye-movements and dreams of the blind. *Quarterly Journal of Experimental Psychology*, 1962, **14**, 183–186.

Berger, R. J., & Oswald, I. Eye movements during active and passive dreams. *Science*, 1962, **137**, 601.

Delange, M., Castan, P., Cadilhac, J., & Passouant, P. Les divers stades du sommeil chez le nouveau-ne et le nourrison. *Revue Neurologique*, 1962, **107**, 271–276.

Dement, W. Dream recall and eye movements during sleep in schizophrenics and normals. *Journal of Nervous and Mental Disease*, 1955, **122**, 263–269.

Dement, W., & Kleitman, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology*, 1957, **53**, 339–346. (a)

Dement, W., & Kleitman, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, 1957, **9**, 673–690. (b)

Dement, W., & Wolpert, E. A. The relation of eye movements, body motility, and external stimuli to dream content. *Journal of Experimental Psychology*, 1958, **55**, 543–553.

Fantz, R. L. Pattern vision in newborn infants. *Science*, 1963, **140**, 296–297.

Fishbein, W., Schaumburg, H., & Weitzman, E. D. Rapid eye movements during sleep in dark-reared kittens. *Journal of Nervous and Mental Disease*, 1966, **143**, 281–283.

Foulkes, D. Dream reports from different stages of sleep. *Journal of Abnormal and Social Psychology*, 1962, **65**, 14–25.

Foulkes, D. NREM-mentation. *Experimental Neurology*, 1967, **4**, 28–38.

Goodenough, D. R., Lewis, H. B., Shapiro, A., Jaret, L., & Sleser, I. Dream reporting following abrupt and gradual awakening from different types of sleep. *Journal of Personality and Social Psychology*, 1965, **2**, 170–179.

Goodenough, D. R., Shapiro, A., Holden, M., & Steinschriber, L. G. A comparison of "dreamers" and "nondreamers": Eye movements, electroencephalograms, and the recall of dreams. *Journal of Abnormal and Social Psychology*, 1959, **59**, 295–302.

Gross, J., Byrne, J., & Fisher, C. Eye movements during emergent Stage I EEG in subjects with lifelong blindness. *Journal of Nervous and Mental Disease*, 1965, **141**, 365–370.

Haynes, H., White, B. L., & Held, R. Visual accommodation in human infants. *Science*, 1965, **148**, 528–530.

Hobson, J. A., Goldfrank, F., & Snyder, F. Respiraand mental activity in sleep. *Journal of Psychiatric Research*, 1965, **3**, 79–90.

Jouvet, M., & Jouvet, D. A study of the neurophysiological mechanisms of dreaming. *Electroencephalography and Clinical Neurophysiology*, 1963, **24**, 133–157.

Karacan, I., Goodenough, D. R., Shapiro, A., & Starker, S. Erection cycle during sleep in relation to dream anxiety. *Archives of General Psychiatry*, 1966, **15**, 183–189.

Ladd, G. T. Contribution to the psychology of visual dreams. *Mind*, 1892, **1**, 299–304.

Luborsky, L., Blinder, B., & Mackworth, N. Eye fixation and recall of pictures as a function of GSR responsivity. *Perceptual and Motor Skills*, 1963, **16** (Monograph Suppl. No. 5), 469–483.

Monad, N., Dreyfus-Brisac, C., Morel-Kahn, F., Pajot, N., & Plussart, E. Les premieres etapes de l'organization du sommeil chez le nouveau-ne pathologique. *Revue Neurologique*, 1964, **110**, 304–305.

Offenkrantz, W., & Wolpert, E. The detection of dreaming in a congenitally blind subject. *Journal of Nervous and Mental Disease*, 1963, **136**, 88–90.

Pivik, T., & Foulkes, D. "Dream deprivation": Effects on dream content. *Science*, 1966, **153**, 1282–1284.

Rechtschaffen, A., Verdone, P., & Wheaton, J. Reports of mental activity during sleep. *Canadian Psychiatric Association Journal*, 1963, **8**, 409–414.

Roffwarg, H. P., Dement, W. C., Muzio, J. N., & Fischer, C. Dream imagery: Relationship to rapid eye movements of sleep. *Archives of General Psychiatry*, 1962, **7**, 235–258.

Roffwarg, H. P., Muzio, J. N., & Dement, W. C. Ontogenetic development of the human sleep-dream cycle. *Science*, 1966, **152**, 604–619.

Spreng, L. F., Johnson, L. C., & Lubin, A. Autonomic correlates of eye movement bursts during stage REM sleep. *Psychophysiology*, 1968, **4**, 311–323.

Verdone, P. Temporal reference of manifest dream content. *Perceptual and Motor Skills*, 1965, **20**, 1253–1268.

Weitzman, E. D., Fishbein, W., & Graziani, L. Auditory evoked response from newborn infants during sleep. *Pediatrics*, 1965, **35**, 458.

Wolpert, E., & Trosman, H. Studies in psychophysiology of dreams: I. Experimental evocation of sequential dream episodes. *Archives of Neurological Psychiatry*, 1958, **79**, 603–606.

(Received December 11, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 2, 159–160

# RESPONSE TO CRONBACH AND FURBY'S "HOW WE SHOULD MEASURE 'CHANGE'—OR SHOULD WE?"[1]

EDWARD F. O'CONNOR, JR.[2]

*Center for the Study of Evaluation, University of California at Los Angeles*

Cronbach and Furby estimated linear combinations of variables by combining least squares estimators of their components. The more conventional least squares estimator of the linear combination as a unit should be used for this purpose. The two procedures produce the same formula for estimating true gain, but for estimating true residual gain the procedure advocated here is more precise.

Lord (1956, 1958, 1963) and McNemar (1958) obtained a least squares estimator of the true gain, $G = Y - X$, using the fallible values, $x$ and $y$. Cronbach and Furby (1970) extended the Lord-McNemar approach by using additional variables, $w$ and $z$, to obtain a more precise estimator:

$$\hat{G} = b_1x + b_2y + b_3w + b_4z, \qquad [1]$$

where $w$ and $z$ represent one or more supplementary measures of the individual's initial and final status. The regression weights, $b_1$, $b_2$, $b_3$, and $b_4$, are chosen to minimize the variance of the errors of estimate, $G - \hat{G}$, and hence they maximize the estimator's precision:

$$\begin{aligned} V &= \Sigma(G - \hat{G})^2 \\ &= \Sigma(G - b_1x - b_2y - b_3w - b_4z)^2. \qquad [2] \end{aligned}$$

This variance is minimized by finding the partial derivatives of $V$ with respect to the regression weights, setting these partial derivatives equal to zero, and solving for the regression weights. This estimator of the true gain can be designated $\hat{G}/xywz$. Similarly, true $X$ and true $Y$ can be estimated using $\hat{X}/xywz$ and $\hat{Y}/xywz$. Cronbach and Furby showed

$$\hat{G}/xywz = \hat{Y}/xywz - \hat{X}/xywz. \qquad [3]$$

Hence the least squares estimator of the true gain can be found either by estimating the true gain as a unit or by estimating true $X$ and true $Y$ separately and subtracting $\hat{X}/xywz$ from $\hat{Y}/xywz$.

Cronbach and Furby (1970) estimated the true residual, $Y \cdot X = Y - B_{YX}X$, as a unit, using a conventional regression equation (Equation 24):

$$\widehat{Y \cdot X} = b_5x + b_6y + b_7w + b_8z. \qquad [4]$$

So far, the argument is sound and internally consistent. A question is to be raised, however, when the residual is formed by partialing out both $X$ and $W$. The true residual in this case is

$$Y \cdot XW = Y - B_XX - B_WW.$$

Instead of estimating $Y \cdot XW$ as a unit, they used $\hat{Y} \cdot \hat{X}\hat{W}$, a combination of estimated true scores:

$$\hat{Y} \cdot \hat{X}\hat{W} = \hat{Y}/xywz - \hat{Y}/\hat{X}\hat{W}, \qquad [5]$$

where $\hat{Y}/\hat{X}\hat{W}$ is the least squares estimator of $Y$ based on $\hat{X}/xywz$ and $\hat{W}/xywz$ (Cronbach & Furby, 1970, p. 77 and Errata).

$\hat{Y} \cdot \hat{X}\hat{W}$ is the least squares estimator of $Y \cdot \hat{X}\hat{W}$, that is, true $Y$ after estimated true $X$ and estimated true $W$ have been partialed out:

$$Y \cdot \hat{X}\hat{W} = Y - B_{\hat{X}}\hat{X} - B_{\hat{W}}\hat{W}. \qquad [6]$$

However an investigator would like to know $Y \cdot XW$, that is, true $Y$ after true $X$ and true $W$ have been partialed out:

$$Y \cdot XW = Y - B_XX - B_WW \qquad [7]$$

$Y \cdot XW$ is the value that he would have if $w$, $x$, and $y$ were perfectly reliable. The least squares estimator of $Y \cdot XW$ is $\widehat{Y \cdot XW}$:

$$\widehat{Y \cdot XW} = b_9x + b_{10}y + b_{11}w + b_{12}z \qquad [8]$$

---

[1] Rodney W. Skager, James L. Ross, Lee J. Cronbach, William H. Angoff, and William B. Schrader made helpful comments on earlier drafts of this paper, although they do not necessarily agree fully with the present version.

[2] Requests for reprints should be sent to Edward F. O'Connor, Jr., who is now at the Division of Analytical Studies and Services, Educational Testing Service, Princeton, New Jersey 08540.

That $\hat{Y}\cdot\hat{X}\hat{W}$ is not equivalent to the more appropriate $\widehat{Y\cdot XW}$ can be shown by a simple example. Assume that $r_{yy} = 1$, $r_{xy} = r_{wy} = 0$, and $r_{zy}$, $r_{zx}$, and $r_{zw}$ are not equal to zero. $\widehat{Y\cdot XW}$ would equal $y$, while $\hat{Y}\cdot\hat{X}\hat{W}$ would equal $y$ minus some function of $x$, $w$, and $z$ (O'Connor, 1972).

It is the writer's understanding that Cronbach and Furby now accept Equations 7 and 8 as the appropriate way to define and estimate true residual gain for individuals (L. J. Cronbach and L. Furby, personal communication, March 1971) Cronbach and Furby suggest that for the convenience of the readers, this paper should mention two points. First, in Formula 14, page 71 of the original article, $\rho_{XX'}$ should replace $\rho_{YY'}$. In addition, one of errata listed on page 218 of the September 1970 *Psychological Bulletin* should be disregarded. The statement that in Equation 21, $\bar{Y}$ should replace $\bar{X}$ inadvertently duplicated a correction that already had been made. The equation is correct as it appears in the original article.

## REFERENCES

Cronbach, L. J., & Furby, L. How we should measure "change"—or should we? *Psychological Bulletin*, 1970, **74**, 68–80. (See also Errata, *Psychological Bulletin*, 1970, **74**, 218)

Lord, F. M. The measurement of growth. *Educational and Psychological Measurement*, 1956, **16**, 421–437. (See also Errata, *Educational and Psychological Measurement*, 1957, **17**, 452.

Lord, F. M. Further problems in the measurement of growth. *Educational and Psychological Measurement*, 1958, **18**, 437–454.

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison: University of Wisconsin Press, 1963.

McNemar, Q. On growth measurement. *Educational and Psychological Measurement*, 1958, **18**, 47–55.

O'Connor, E. F. Extending classical test theory to the measurement of change. *Review of Educational Research*, 1972, **42**, 73–97.

(Received January 18, 1971)

VOL. 78, NO. 3 SEPTEMBER 1972

Psychological Bulletin

Diary No. 733(3)
Date 1.2
[illegible] Lib
Bureau Ed. & Psy. Research.

# BIRTH ORDER EFFECTS:
## NOT HERE, NOT NOW!

CARMI SCHOOLER[1]

*National Institute of Mental Health, Bethesda, Maryland*

A review of birth order effects, which takes into account previously reported and unreported data and recently hypothesized biases arising from long-term population trends, reveals almost no reliable evidence for birth order effects among men living in the United States in the middle 1960s, and only a marginal increase in such evidence when such restrictions on time, place, and sex are removed. This conclusion is based on data from both psychiatric and normal populations examined by comparing (*a*) the prevalence of particular birth ranks in relevant populations, (*b*) the characteristics of individuals of different birth ranks, (*c*) the parents' reports of their treatment of children of different birth ranks.

At this point in time a good case could be made that still another review of the effects of birth order would be, to say the least, redundant. The present article, in fact, is considerably indebted to reviews and bibliographies of the subject by Chen and Cobb (1960), Schachter (1959, 1963), Clausen (1965), Warren (1966), Altus (1967), and Miley (1969). However, a new evaluation is made necessary both by the recent availability of a wide range of previously unreported data showing essentially negative findings, and by the apparent validity of recently raised hypotheses about potential sources of bias arising from long-term population trends. The general conclusion reached is that for people living in the United States in the middle 1960s there is almost no evidence of any birth order effects among men, and that these effects increase only marginally when restrictions on time, place, and sex are removed. This conclusion is based on data from both psychiatric and normal populations that are examined in three ways. One involves the comparison of the prevalence of particular birth ranks in theoretically relevant subpopulations; the second involves the direct comparison of relevant characteristics of individuals of different birth ranks; while the third evaluates a potentially major cause of birth rank differences by comparing parents' reports of their treatment of children of different birth ranks.

## PREVALENCE OF DIFFERENT BIRTH RANKS

### *Psychiatric Populations*

A high proportion of the studies of the prevalence of different birth ranks among psychiatric populations have centered on schizophrenia. Recently two papers appeared in the same issue of the *British Journal of Psychiatry* which, unbeknown to their authors, provided a striking confirmation of each other. In one, Erlenmeyer-Kimling, van den Bosch, and Denham (1969)—after reviewing the data from four large studies of birth order and schizophrenia—observed:

> one feature . . . does appear consistently . . . in the four Western samples of both sexes. Patients from sibship sizes 2–3 tend to be earlier-born than patients from sizes 5 or greater in the same sample . . . [p. 669].

In the other article, Hare and Price (1969) pointed out, in discussing biases in birth rank

---

[1] Requests for reprints should be sent to Carmi Schooler, National Institute of Mental Health, Laboratory of Socio-Environmental Studies, Building 10, Room 3D41, Bethesda, Maryland 20014.

© 1972 by the American Psychological Association, Inc.

created by changes in family size and birth rate:

> Changes in the birth rate of a population may be the result of changes in the number of families started or of changes in the size of completed families . . . An increase in the number of families started will result in an over-representation of early birth ranks for every sibship size. A decrease in family size will result in an over-representation of early birth ranks in small sibships [p. 647].

In other words, (*a*) in a population where there has been a long-term trend toward a greater number of new families being started each year than in the previous year, there should be more firstborn than last-born individuals; (*b*) in a population where the average family size is decreasing so that the number of large families (i.e., more than five children) completed each year is smaller than the number of large families completed the year before, there should be more last-born than firstborn individuals from large families. Hare and Price (1969) supported their predictions both by their analyses of the birth ranks of all Scottish children who were 11 years old in 1947, and of a random sample taken by Kohn and Schooler in 1964 of 3,101 employed males in the United States. Given these validations of Price and Hare's predictions on large samples from different populations, it is not surprising that Erlenmeyer-Kimling et al. (1969), although unaware of the cause found that

> a trend toward later birth orders with increasing sibship size is about as likely to be found in samples of nonschizophrenic subjects as in samples of schizophrenic patients. Such a trend is the sole aspect of the birth order data . . . that fairly consistently describes the schizophrenic patients [p. 670].

The present article provides qualified support for the obvious hypothesis that all of the previously reported findings of relationships between birth order and schizophrenia stem from the types of biases described by Price and Hare (1969) that are the results of long-range trends toward an increase in the number of families started and a decrease in family size. It does this by supplementing the data presented by Erlenmeyer-Kimling et al. with a reanalysis of data from other previously published studies, and with presentation of new data made available from several sources. The relevant data, the sources of the samples, and the nature of the populations studies are presented in Tables 1 and 2. These tables contain all studies that could be found which were based on firm sampling criteria, and where it was possible families were classified into small (2–4 children) and large (5+ children) groups.

As had been pointed out by several authors (Chen & Cobb, 1960; Gregory, 1958; Schooler, 1961), other things being equal:

> in any population sample from which "only" children are excluded, the ratio of the number of eldest to the number of youngest children should be precisely one, subject only to sampling variation. This is because each sibship of size two or greater has both an eldest and a youngest child [Chen & Cobb, 1960, p. 552].

That other things are not equal is shown by the fact that in 10 out of 12 studies, the proportion of firstborn to last-born individuals is greater in small families than in large families. This is also the case in 17 out of 20 of the analyses that could be performed in the 10 studies where the data allowed separate analyses by sex. This same finding also holds in two occasions where the break between small and large families was made at four rather than at five children (Schooler, 1961; Lee, Pollin, Ladusky, & Perlin, cited in Schooler, 1961), and in one study where the data were presented in terms of sibship halves (Farina, Barry, & Garmezy, 1963).[2] Table 2 —which contains data from non-American sources—reveals that trends in Japan, England, and Canada also follow the same general pattern as the United States. The only findings counter to these general trends are reported by Sundararaj and Rao (1966) for schizophrenic patients in nonindustrial Bangalore, India.

---

[2] A similar finding occurs in the data collected by Solomon and Nuttall (1967). The proportion of firstborns to last-borns from families of 2–4 sibs is 4.158 and from families of 5+ sibs is 2.667. Their sample, however, was a selective one. They noted, "We made a deliberate effort to secure a high proportion of younger, first admission patients . . ." and two-thirds of their patients came from hospitals which "have a disproportionate number of upper-social-class patients." As is seen in a later section of the paper, such selectivity in terms of class and age would bias the results toward an overrepresentation of firstborn sibs.

TABLE 1

DATA, SOURCES OF SAMPLES, AND NATURE OF THE POPULATIONS: AMERICAN STUDIES

| Source | Years sampled | Sampling procedure | Birth order | Family size | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Total | | Male | | Female | |
| | | | | 2–4 | 5+ | 2–4 | 5+ | 2–4 | 5+ |
| Schooler (1964)[a] | 1942–49 | All patients admitted to Spring Grove Hospital who were given schizophrenic diagnosis and no organic involvement. | Firstborn<br>Last born | 120<br>148 | 55<br>67 | 59<br>51 | 28<br>19 | 61<br>97 | 27<br>48 |
| | | | Firstborn / Last born | .811 | .821 | 1.157 | 1.474 | .629 | .563 |
| Barry & Barry (1967) | 1942–62 | All patients between 18–59 years of age admitted to Western Psychiatric Institute and Clinic given final diagnosis of schizophrenia. | Firstborn<br>Last born | 195<br>168 | 50<br>97 | 70<br>54 | 15<br>39 | 125<br>114 | 35<br>58 |
| | | | Firstborn / Last born | 1.161 | .515 | 1.296 | .385 | 1.10 | .603 |
| Malzberg (1940) | 1930–35 | First admissions with dementia praecox at Manhattan State Hospital (excluding those with half siblings and illegitimate). | Firstborn<br>Last born | 87<br>88 | 39<br>57 | —<br>— | —<br>— | —<br>— | —<br>— |
| | | | Firstborn / Last born | .989 | .684 | — | — | — | — |
| Erlenmeyer-Kimling et al. (1969) | 1934–36 | Consecutive admissions to 11 state mental hospitals in New York with schizophrenic diagnosis verified by staff psychiatrists involved in survey. | Firstborn<br>Lastborn | 121<br>89 | 42<br>51 | 70<br>36 | 24<br>30 | 51<br>53 | 18<br>21 |
| | | | Firstborn / Last born | 1.360 | .824 | 1.944 | .800 | .962 | .857 |
| Erlenmeyer-Kimling et al. (1969) | 1964 | As above. | Firstborn<br>Last born | 143<br>107 | 34<br>42 | 83<br>49 | 14<br>18 | 60<br>58 | 20<br>24 |
| | | | Firstborn / Last born | 1.336 | .810 | 1.694 | .777 | 1.034 | .833 |
| Eight-hospital collaborative study (1966) | 1963–64 | Newly admitted schizophrenics in 8 eastern hospitals between 1963–64, approximately 60 at each hospital, equally divided between male and female. | Firstborn<br>Last born | 93<br>73 | 25<br>37 | 48<br>36 | 10<br>18 | 45<br>37 | 15<br>19 |
| | | | Firstborn / Last born | 1.274 | .676 | 1.333 | .555 | 1.216 | .789 |
| Total | | | Firstborn<br>Last born | 759<br>673 | 245<br>351 | 330<br>226 | 91<br>124 | 342<br>359 | 115<br>170 |
| | | | Firstborn / Last born | 1.127 | .698 | 1.460 | .734 | .953 | .676 |

[a] Family size classifications in the original article were 2–3 and 4+; the data were regrouped for the present table.

It is possible, however, that another—albeit much less powerful—birth order effect exists. If we examine the total numbers of firstborn and last-born women in the American studies (Table 1) in which large families are defined as those consisting of five or more children, the proportion of firstborn to last-born women in small families (four or less) is less than one (.953) and significantly less than the proportion of firstborn to last-born men in families of the same size (1.460) ($\chi^2 = 13.49$, $df = 1$, $p < .001$).[3] It is possible, then, that in this population last-born women from small families are in fact to some slight extent overrepresented among hospitalized schizophrenics. Interestingly, the Japanese sample gathered in 1958 shows a similar significant trend.

The data on the relationship between birth order and hospitalization for psychiatric reasons other than schizophrenia are even less impressive than those for schizophrenia. Gregory (1959) in a study of 1,000 consecutive admissions to a Canadian mental hospital between 1954 and 1958 found no significant differences in the birth ranks of

---

[3] On the other hand, the English sample collected by Price and Hare (1969) shows more last-born than firstborn males from families of less than four, although the proportion of firstborn to last born among men is not significantly less than that among women.

TABLE 2

DATA, SOURCES OF SAMPLES, AND NATURE OF THE POPULATIONS: NON-AMERICAN STUDIES

| Source | Year sampled | Sampling procedure | Birth order | Family size: Total 2–4 | Total 5+ | Male 2–4 | Male 5+ | Female 2–4 | Female 5+ |
|---|---|---|---|---|---|---|---|---|---|
| | | | Canadian | | | | | | |
| Gregory (1959) | 1954–58 | 1,000 consecutive patients born 1906–35 inclusive, admitted at least once to Ontario Hospital, London, Canada. | Firstborn | 55 | 28 | — | — | — | — |
| | | | Last born | 64.5 | 38 | — | — | — | — |
| | | | Firstborn/Last born | .853 | .734 | — | — | — | — |
| | | | English | | | | | | |
| Granville-Grossman (1966) | 1942–62 | All inpatients over 16 years of age discharged from Maudsley and Bethlem Royal Hospitals diagnosed schizophrenic. | Firstborn | 229.5 | 48.5 | 93 | 17.5 | 136.5 | 31 |
| | | | Last born | 241 | 71 | 117 | 40.0 | 124 | 31 |
| | | | Firstborn/Last born | .952 | .683 | .795 | .438 | 1.101 | 1.00 |
| Hare & Price (1966)[a] | 1958–66 | All cases attending Bethlem and Maudsley Hospitals and discharged during period sampled. | Firstborn | 368 | 74 | 185 | 40 | 183 | 34 |
| | | | Last born | 384 | 112 | 211 | 56 | 173 | 55 |
| | | | Firstborn/Last born | .958 | .661 | .877 | .714 | 1.058 | .618 |
| | | | Japanese | | | | | | |
| Caudill & Schooler (1969) | 1958 | Consecutive admissions to 4 Japanese mental hospitals given schizophrenic diagnoses. | Firstborn | 39 | 11 | 21 | 7 | 18 | 4 |
| | | | Last born | 32 | 21 | 14 | 10 | 18 | 11 |
| | | | Firstborn/Last born | 1.219 | .524 | 1.50 | .70 | 1.00 | .364 |
| Unpublished data gathered by Caudill[b] | 1964 | Consecutive admissions to 5 Tokyo mental hospitals given schizophrenic diagnoses. | Firstborn | 86 | 21 | 48 | 9 | 38 | 12 |
| | | | Last born | 55 | 31 | 29 | 16 | 26 | 15 |
| | | | Firstborn/Last born | 1.564 | .677 | 1.655 | .563 | 1.462 | .800 |
| | | | Indian | | | | | | |
| N. Sundararaj & B. S. S. R. Rao (1966) | 1963–65 | Records of 1,034 patients admitted to Bangalore Mental Hospitals selected on the basis of completeness of relevant information. | Firstborn | 110 | 140 | 82 | 86 | 28 | 54 |
| | | | Last born | 123 | 72 | 85 | 50 | 38 | 22 |
| | | | Firstborn/Last born | .894 | 1.944 | .965 | 1.72 | .737 | 2.455 |

[a] J. S. Price kindly provided the data on birth rank and sibship size for schizophrenics for both the total sample and each sex separately. It should be noted that there is some overlap between the two English samples.

[b] Personal communication; a debt of gratitude is owed to the late William Caudill for providing this data.

manic-depressive, psychoneurotic, pathological personality, and alcoholic patients. Hare and Price (1969) examined the birth order of all the cases discharged from the Bethlem and Maudsley Hospitals between 1958 and 1966 and found that for five major nonschizophrenic diagnostic groups, each of more than 1,800 patients, the distributions of birth ranks are not appreciably different, and are in accordance with theoretical predictions based on the number of marriages and of births in the population of England and Wales.

Alcoholism is the behavioral pathology that has been most frequently linked in the past to birth rank differences. Schachter (1959) found, in his original monograph on the psychology of affiliation after reanalyzing data provided by Bakan (1949), that alcoholics were overrepresented among later-born persons. Gregory (1965), however, has shown through analysis of his own data, further analyses of Bakan's data, and data provided by Smart (1963a, 1963b) that if there are any birth order differences among alcoholics they are due to an overrepresentation of lastborns from large families, a finding that we now expect on the basis of the Price and Hare (1969) hypothesis. Chen and Cobb (1960) summarized data from seven other studies examining the relationship between birth order and alcoholism: the ratio of firstborn to last-born in these seven studies ranges from 1.7 to .6, with no significant overall trend observable.

These same authors, in discussing supposed birth order differences in two other kinds of pathological behavior—the excess of juvenile delinquency (Glueck & Glueck, 1950; Sletto, 1934) and anorexia nervosa (Cobb, 1950) in the "intermediate position"—cautioned:

> It is not safe to conclude that this is entirely due to an effect of intermediate birth order, as is suggested by some authors, . . . for it might equally be an effect of social class acting through family size [p. 556].

Rheumatoid arthritis, the only physical illness with a potential psychiatric basis for which Chen and Cobb provide data both on birth rank and sibship size, also follows the pattern described by Price and Hare of being significantly more prevalent among later-born than among earlier-born siblings, occurring significantly more frequently among those from larger families. Finally, a reexamination of previously reported associations between birth order and homosexuality in both males (Slater, 1962) and females (Gundlach & Riess, 1967) reveals the same pattern of there being more firstborn individuals from small families and more later borns from large families.

In summary, it would seem that with the possible exception of a slight surplus of last borns from small families among hospitalized American female schizophrenics, there do not seem to be any differences in the prevalence of different birth ranks among psychiatrically abnormal populations that are not explained parsimoniously by the Price and Hare (1969) hypothesis.

### *Normal Populations*

As noted, Price and Hare (1969) supported their hypothesis in part with data from a national study of the social psychological effects of occupational experiences by Kohn and Schooler (1969), which involves a representative sample of 3,101 males employed in the United States in 1964. Further analyses of that National Occupation Study sample reveal that family size is inversely related to the level of father's education ($r = -.255$), and the proportional excess of last borns from large families is equally great among those from working-class and middle-class origins. One would expect an overrepresentation of last borns among the respondents from working-class families which tend to be large, and correspondingly, a proportionate overrepresentation of firstborn children from middle-class families.[4] As is shown in Table 3, this is the case. It is therefore not surprising that since family size has not been controlled in many of them, a long series of studies have found firstborn individuals to be overrepresented among populations in which one would normally expect to find a high proportion of people with middle-class origins: eminent men (Ellis, 1904), Rhodes scholars (Apperly, 1939), people in *Who's Who* (Jones, 1954), university professors (Gini, 1915), psychologists (Roe, 1953), and, most importantly, college students.

The importance of this latter finding is stressed by Schachter (1963) in assessing the nature of the effect of birth order on eminence and achievement. He concluded:

> The repeated findings of a surplus of first-borns among eminent scholars appears to have nothing to do with any direct relationship of birth order to eminence but is simply a reflection of the fact that scholars, eminent or not, derive from a college popu-

---

[4] There is also some evidence of a surplus of firstborns among those whose fathers completed high school among both small and large families, which is probably due to the increasing number of people who graduate from high school. This would result in a trend toward a greater number of families being started by high school graduates and hence toward a greater number of firstborns from such origins.

TABLE 3

EFFECTS OF SOCIAL CLASS DIFFERENCES IN FAMILY SIZE ON THE BIRTH RANK OF THEIR MEMBERS IN THE NATIONAL OCCUPATION STUDY

| Level of father's education | $N$ | % from large[a] families | Relative excess of last borns from large families[b] | Proportion of firstborns to last borns |
|---|---|---|---|---|
| Not high school graduate | 819 | 34.5 | 1.29 | 1.17 |
| High school graduate | 323 | 14.9 | 1.33 | 1.78 |

Note.—$\chi^2 = 9.259$, $df = 1$, $p < .01$.

[a] Families with 5 or more.

[b] Proportion of last borns to firstborns of large families divided by proportion of last borns to firstborns of small families.

> lation in which first-borns are in marked surplus [p. 768]. ✓

His analysis of birth ranks of 4,000 introductory psychology students at the University of Minnesota in 1959 and 1961, which takes into account both the effects of family size and the proportion of firstborns to total births in the year the subjects were born, shows a predominance of firstborn individuals. ✓ Bayer (1966), however, in a later study that used a sample of 50,000 high school students, and then examined the percentage of students in senior college one year after high school graduation, concluded:

> last born are about as likely to go to college as are first-born from the same family size and family socioeconomic position. . . . Within each socioeconomic class and each family size . . . those of an intermediate ordinal position are consistently the least likely to attend college [p. 484]. ✓

When the relationship between birth order and college attendance among the employed males in the National Occupation Study[5] is examined without controlling on family size, there is no difference between firstborn and last-born individuals in the proportion going to college, although middle-born individuals—a higher proportion of whom come from larger families—are significantly underrepresented. When family size is controlled there are no significant differences among any of the birth ranks, although there is a tendency toward overrepresentation of firstborns from small families. However, when father's education is controlled even this tendency disappears (Table 4). It would thus appear fair to say that the overrepresentation of firstborns found by others among such middle-class populations as college students is due to differences among the social classes in family size and birth rate trends and that, in fact, social class of origin should be controlled whenever the effects of birth order are examined.

There are, however, several birth order findings that involve achievements somewhat beyond the range of average middle-class individuals, where some attempt has been made to control for the effects of family size. The most recent and impressive of these is the study by R. C. Nichols on birth order and intelligence (cited by Altus, 1967) of 618 National Merit Scholarship finalists whose test scores imply an aptitude "at least in the top .5% of the general population. . . ." These data indicate that within each family size firstborn individuals are overrepresented. However, in evaluating this overrepresentation of firstborns, it should be kept in mind that if these students were 17 years old when they took the test they were born in 1947, which was the peak period for the starting of new families in our country; fully 43% of the babies born that year were firstborn.[6] If the middle class was disproportionately overrepresented among those starting new families during this period, Nichols' findings would cease to be impressive. This, in fact, would seem to be the case, since there is strong evidence (Grabill, Kiser, & Whelpton, 1958) that

> the 1940–1950 period was marked by important increases in fertility among women under 40 years of age and the magnitude of the percent increase differed systematically by occupation group of the husband. . . . the 1940–1950 increase in the cumulative fertility rate ranged from 53 percent for wives

---

[5] The data used in this analysis were collected as part of the study conducted by Kohn and Schooler (1969) on the social psychological consequences of occupational experience. Interviews were conducted with 3,101 men representative of all men throughout the United States employed in civilian occupations. These interviews were conducted for us by the National Opinion Research Center in the spring and summer of 1964.

[6] The percentages of firstborn in 1946 and 1948 were also very high, being 40% and 39%, respectively (Department of Health, Education, and Welfare, 1959).

TABLE 4

Relationship between Birth Order, College Attendance, and Father's Education[a]

| Respondent's father's social class and education | Birth order | Family size | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2–4 | | | 5+ | | |
| | | Total $N$ | Respondent's education | | Total $N$ | Respondent's education | |
| | | | % <College | % College | | % <College | % College |
| Lower class; < high school | Firstborn | 301 | 69.4 | 30.6 | 141 | 84.4 | 15.6 |
| | Middle born | 243 | 71.6 | 28.4 | 669 | 86.2 | 13.8 |
| | Last born | 236 | 70.8 | 29.2 | 142 | 82.4 | 17.6 |
| | | | $\chi^2 = .3146$<br>$df = 2$; *ns* | | | $\chi^2 = 1.5285$<br>$df = 2$; *ns* | |
| Upper class; high school and higher | Firstborn | 179 | 62.0 | 38.0 | 28 | 64.3 | 35.7 |
| | Middle born | 99 | 53.5 | 46.5 | 116 | 63.8 | 36.2 |
| | Last born | 97 | 63.9 | 36.1 | 20 | 45.0 | 55.0 |
| | | | $\chi^2 = 2.6407$<br>$df = 2$; *ns* | | | $\chi^2 = 2.6507$<br>$df = 2$; *ns* | |

[a] Limited to those cases where father's education is known.

> of professional men to 19 percent for wives of laborers . . . [pp. 384–385].

A study of birth order and scientific attainment by Datta (1967), using a somewhat similar population—Westinghouse Talent Search entrants—concluded:

> Among students of high scientific aptitude first-borns appear to be overrepresented among young men of unusual early scientific attainment but not to any greater extent than among young men of lesser attainment [p. 158].

Since these data are based on young scientists who competed in the 1963 Westinghouse Talent Search and who were born in the middle 1940s, it is not surprising that firstborns are overrepresented among members of this age group—whatever their degree of scientific attainment. Finally, in describing the characteristics of stripteasers, Skipper and McCaghy (1970) noted that of the 35 strippers in their sample, 31 (89%) were firstborn, and of these just 5 were in the only-child category. They also noted that 60% of the total sample was between the ages of 20 and 30. It is probably safe to assume that some of the excess representation of firstborns among stripteasers in this study occurred because some members of the postwar baby boom did not become scientists when they grew up.

## Comparisons of Characteristics of Individuals of Different Birth Ranks

Although the above evidence casts strong doubt on hypotheses predicting differences in the incidence of different birth ranks in various populations (e.g., schizophrenics, college students), the possibility still exists that birth order does have an effect on the characteristics of individuals within such populations. Firstborn schizophrenics may behave differently from last-born schizophrenics, and firstborn college students differently from last-born ones.[7]

---

[7] Studies in which only-born children were included in the firstborn category were excluded from the following review. This was done for both theoretical and empirical reasons. Only-born children differ from firstborns from families of two or more because they have never had the experience of having sibs. Looked at another way, only-born children can be said to be as similar to last-born sibs, who also have never had a younger sib, as they are to firstborn sibs, with whom they share the characteristic of never having had an older sib. It also seems plausible to suspect that parents who only have one child have distinctly different characteristics from those who have more than one child. That these theoretical considerations can have empirical consequences is shown by the fact that in studies where the two types have been compared, directly significant differences have emerged. Thus in a partial replication of Schachter's

*Psychiatric Populations*

Most of the older data on the effects of birth order among pathological populations indicate that birth order tends to have stronger effects among women than among men, and that later-born women tend to be sicker than earlier-born women. Schooler (1964), in a study based on records of patients hospitalized between 1942 and 1949, found that "Compared to first-born patients, last-born hospitalized female schizophrenics apparently have a significantly lower degree of social competence [p. 577]" and are more likely to have hallucinations, feelings of depersonalization, and suicidal tendencies. No significant differences were found among men.

Several studies using behaviorally based measures also indicate that later-born hospitalized female schizophrenics are also less sociable and possibly sicker than earlier-born ones. Schooler and Scarr (1962) found a biserial correlation of .39 ($p < .05$, two-tailed) between sociability while awaiting testing and birth order, subjects in the first half of their family sibships being more sociable. In a second study which involved structured observations of patient behavior in the ward setting, Schooler and Parkel (1966) concluded:

> female schizophrenics born in the last half of their sibling group were significantly more likely to be found against the wall, less likely to be depressed, and less likely to engage in either verbal social or nonverbal social behavior than were those born in the first half of their sibling group [p. 73].

If likelihood of recovery can be considered as a characteristic of an individual's illness, data presented by Farina et al. (1963), based on schizophrenics admitted to North Carolina State hospitals between 1949 and 1954, suggest that patients with many older siblings are less likely to recover than are those with few older siblings. Furthermore, in their sample at least, the female patients had more older siblings than did the male patients. However, the authors, being aware of the relationship between family size and social class in their sample, noted that the relationship between birth order and recovery may be a spurious one, the actual relationship being between social class and recovery.

Analyses of the effects of birth order (Caudill & Schooler, 1969) on the recorded symptoms of a non-American sample—658 nonorganic patients admitted in 1958 to four Tokyo mental hospitals—revealed that, in terms of birth rank, eldest sons were likely both to be more depressed and to act out more than were other sons, while eldest daughters never showed more symptoms than did women of other birth ranks.

Despite the somewhat consistent finding in both the earlier Japanese and American studies—that elder daughters among hospitalized female schizophrenics apparently show a less severe pattern of symptoms—our analysis of the data from the most recent and most reliable American study of newly admitted hospitalized schizophrenic patients shows no significant birth order effects among women or men. These data, which were collected for an eight-hospital collaborative psychopharmacological study conducted by the Psychopharmacology Research Branch of the National Institute of Mental Health,[8] are based on a sample of 480 patients which was designed to provide a cross section of schizophrenics admitted to mental hospitals in the eastern United States in 1964. The symptom data provided by this study are based on psychiatric interviews administered by trained personnel who assessed the patients' responses with the Lorr Inpatient Multidimensional Psychiatric Scale (Lorr & Klett, 1965). Once family size was controlled, no significant birth order effects were found for any of the symptom factors. Neither were there any significant birth order effects in ward behavior as rated by nurses on a ward behavior rating scale, on global assessment of the degree of mental illness made by either doctors or nurses, on the patients' prognosis,

---

study of affiliation under stress, Schooler and Scarr (1962) found that among the hospitalized female schizophrenics tested, only-born children were the least sociable group and firstborns the most sociable.

[8] *Eight hospital collaborative study of the effects of phenothiazines on acute schizophrenia.* Washington, D. C.: National Institute of Mental Health, 1966. (The author is indebted to Solomon Goldberg of the Psychopharmacology Research Branch for providing the data from this study, which was supported by NIMH Grants MHO4661, MHO4663, MHO4667, MHO4673, MHO4674, MHO4675, MHO4679, MHO-4803.)

nor on the actual outcome of their stay in the hospital. This total absence of findings for newly admitted patients of either sex in the United States in the mid-1960s strongly suggests that in this country, at that point in time, birth order was not affecting the likelihood of hospitalization, symptomatology at the time of hospitalization, or outcome once hospitalized.

### *Normal Populations*

In an extensive review of the literature Warren (1966) concluded:

> After considerable attention to birth order in its relationship with social behavior, only two or three conclusions have found substantial support. . . . Firstborn of both sexes are more susceptible to social pressure and are more dependent than later born. Firstborn women, when apprehensive, desire the company of others more strongly than do later-born women [p. 47].

Doubts about the certainty of even these conclusions are raised by two major subsequent studies, one by Sampson and Hancock (1967) and the other by Rosenfeld (1966).

Samson and Hancock's study involved 251 high school students who were tested experimentally for conformity to group norms and given the Mandler-Sarason Test Anxiety Scale and the Edwards Personal Preference Schedule, which was scored for Need Achievement, Need Affiliation, and Need Autonomy. They found that firstborns had more Need Autonomy than later borns, in contradiction to Warren's (1966) conclusion that firstborns are generally more dependent. Predicted differences between firstborns and later borns in conformity behavior emerged only after examination of an insignificant interaction ($p < .10$), which showed that firstborn males conformed more to social pressure than did later-born males. There were no birth order effects at all among females, thus contradicting Warren's conclusion that conformity marks firstborns of both sexes. In areas in which Warren did not find any definite birth order differences, but which were considered important by other investigators, Samson and Hancock found no birth order differences at all for Test Anxiety and Need Affiliation. The firstborns in their study did show greater Need Achievement than did later borns, but since the authors noted that there were differences among the cells in the level of the subjects' fathers' education this finding may just reflect the tendency of middle-class individuals to be firstborn and to have high Need Achievement.

The article by Rosenfeld (1966) casts even more doubt on the existence of any general birth order effects. It is based on five studies, each involving over 80 subjects. He noted:

> Preparation of the present report was stimulated by two published studies of the relationship between ordinal position and projectively assessed need for affiliation, each of which claimed support for incompatible hypotheses [p. 467].

He found no differences in Need Affiliation among birth ranks as measured by the Thematic Apperception Test, either in neutral conditions or under conditions of arousal, generally negative results in terms of Need Achievement, and a mild trend across the studies for later borns to show more Test Anxiety than firstborns, a tendency which does not occur in Sampson and Hancock's (1967) studies. Rosenfeld (1966) concluded:

> The many negative and paradoxical findings presented . . . in this report certainly call for qualification of the general propositions that first-born persons surpass later-born persons in affiliative motivation and achievement motivation. . . . The repeated contradictions to published findings, in the present set of studies, may also have serious implications for the tendency for unreplicated positive results to become published, while negative results tend to remain private [p. 478].

A review of papers, too recent to be included in the Warren (1966) summary, shows some support of previous birth order hypotheses. Thus Moran (1967), in a study in which the Marlowe-Crowne Social Desirability scale was used to measure need for approval among 348 introductory psychology students, found that firstborn subjects showed higher need for approval. A study of personality, intelligence, and school achievement among 4,000 English primary school children by Eysenck and Cookson (1969) provided some confirmation of the hypothesized greater scholastic ability of firstborn children. When the effects of family size were controlled, firstborn children did better than later-born children in English and mathematics achievement tests. However, the authors remained puzzled by this finding since there were no significant birth rank differences in reading and verbal reasoning ability as well as in the nine other dependent variables measured, nearly all of which have been hypothesized to be related to birth order: grammar school entrance, extraversion, neuroticism, and lying as measured by the Eysenck Junior Personality Inventory, parental interest and teachers' ratings of emotional stability, perseverance, sociability, and impulsiveness.

Similarly, the preponderance of recent American studies comparing the experimental or psychological test performance of individuals of different birth rank, reported during the years since Warren's (1966) review was written, shows either no results or results contrary to previous findings. Thus, Rees and Palmer (1970) in analyzing mental test performance of over 1,500 children who were the subjects of longitudinal studies, found "Once partial correlations controlling for SES [socioeconomic status] were computed, neither ordinal position nor family size was related to test performance [p. 44]." Alexander (1966), in a study of the relationship between birth order and sociometric studies of 1,410 male high school seniors, found results contradictory to Schachter's (1963) findings that firstborns choose relatively less popular persons for friends and receive more sociometric choices and more reciprocations of choices than later borns. Rhine (1966), testing 128 grade-school girls with experimentally based measures of conformity in different arousal conditions, found no significant birth order differences in either conformity or arousal. McKeithen (1965), after giving the Edwards Personal Preference Inventory to 495 women undergraduates and looking for birth order differences in Need Achievement, Need Autonomy, and Need Affiliation, etc., found "despite the large number of predictions made, very few of the differences found proved to be statistically reliable [p. 4585]." Farley (1967), giving the Mandler-Sarason Test Anxiety Scale and the Need Achievement Scale of the Edwards Personal Preference Schedule to 49 male university students, found no significant differences between birth ranks. Wolkon and Levinger (1965), citing data from three studies where respondents were given the Edwards Personal Preference Schedule, found "None of the 12 separate *t* tests contrasting scores of first- and later-borns was statistically significant. In fact, fewer than half were in the predicted direction [p. 74]." Masling (1965), in a sample of 50 male and 195 female college freshmen, could find no difference in the affiliative behavior of firstborn and later borns as reflected in their requests for roommates. In another sample of 153 athletes he could find no relation between birth rank and effectiveness, that is, being selected for the varsity in group or individual sports, although such a relationship could well have been predicted on the basis of Schachter's (1959) original findings that effective fighter pilots are more likely to be later born.

Similarly, negative results occur in the analysis of birth order effects in the 3,101-subject representative sample of the previously cited National Occupation Study. Birth order differences were examined by dividing the subjects into the following five groups: firstborn; first half and not firstborn; middle children in odd-sized families; last half and not last born; and last born. In all analyses, effects of family size were controlled through covariance and only-born children were excluded. The dependent variables used were based on factor analyses of the subjects' responses in an interview which, on the average, lasted $2\frac{1}{2}$ hours. The 21 dependent variables listed in Figure 1 covered a wide range of values and orientations to self and others, as well as several centrally important measures of psychological and intellectual functioning. No significant differences were found. This lack of findings is, in fact, a record of sorts since, given the large number of cases and the relative discriminatory power of the dependent variables, nearly all other analyses with this body of data have produced some pattern of significant results regardless of whether the independent variables represented occupational conditions, background factors, or even such aspects of family structure as family size, intactness of family, and having a mother who worked (Kohn, 1969; Kohn & Schooler, 1969).

Since the interviewers gathered data not only on the respondent's occupation but also on the occupations of his siblings, it also

INTELLECTUAL FUNCTIONING
1. Ideational flexibility
2. Perceptual flexibility

SELF-CONCEPTION
3. Self-confidence
4. Self-deprecation
5. Sense of control over fate
6. Anxiety
7. Believes ideas conform to those of groups to which one belongs
8. Compulsiveness

SOCIAL ORIENTATIONS
9. Authoritarian–conservatism
10. Pragmatic morality
11. Trustfulness
12. Receptiveness to change

VALUES FOR SELF
13. Self-judging values
14. Self-oriented values

LIFE GOALS
15. Pleasure for self rather than truth and benefiting others
16. Success rather than aesthetic satisfaction

JUDGMENTS ABOUT WORK
17. Emphasize intrinsic qualities of job
18. Emphasize extrinsic benefits of job

USE OF LEISURE TIME
19. Social
20. Family centered
21. Intellectually demanding

FIG. 1. Dependent variables for which no significant birth order effects were found in the representative sample of 3,101 employed men in the National Occupation Study.

proved possible to see whether there were any birth order differences in occupational mobility. This was done by assigning each respondent a score representing the difference between the status of his occupation, as measured by the Hollingshead ranking of occupations (Hollingshead & Redlich, 1958), and the average occupational status of his brothers. Analysis of these scores for different birth ranks showed no significant differences. This lack of results, given the size and representativeness of the sample, casts strong doubt on the existence of any effect of birth order on the level of occupational attainment in the general population.

### Differences in Parental Approach to Children of Different Birth Ranks

Although they do not make much of it, Sears, Maccoby, and Levin (1957) found some evidence of significant differences in the approach parents took to children of different birth ranks. Their evidence resulted from an extensive analysis of the self-reported child-rearing patterns of the parents of American-born children attending public kindergarten in a New England suburb during 1951 and 1952. Their findings are indicative of a general tendency toward preferential treatment of firstborn children, parents being more likely to be pleased about a first than later pregnancy, and continuing to show more interest in and leniency toward their firstborn in both weaning and discipline.

On the other hand, analysis by both Kohn and Schooler of data collected by Kohn in 1956 on a sample of 174 middle-class and 165 working-class mothers of 10-year-old children revealed no significant patterns of birth order effects either in terms of child-rearing values or in reported reactions to children's misbehavior (Kohn, 1969; Schooler, 1964). Furthermore, analysis of the responses of 1,516 men in the 1964 National Occupation Study sample who had children between 3 and 15 years of age also failed to reveal any differences related to birth rank of the child in child-rearing values or in the reported display of affection or anger. Thus, it would seem quite likely that if there were any substantial differences in American parents' approaches to children of different birth ranks in the early 1950s, these differences disappeared by the mid-1960s.

### Discussion

The present article has attempted to survey the relationships between birth order and various normal and abnormal psychological characteristics in three ways: (*a*) looking at the prevalence of different birth ranks in the relevant populations, (*b*) comparing various the characteristics of individuals of known birth rank, and (*c*) examining parents' reports of their treatment of children of different birth ranks. Each of these approaches has produced essentially negative results.

In terms of psychiatric abnormalities, there appears to be some evidence of a slight surplus of last-born individuals from small families among American women hospitalized for schizophrenia. With the possible exception of such a minor trend, once the effects described by Price and Hare of long-term trends in the numbers and sizes of families started are taken into account, there does not seem to be any reliable evidence of a relationship between birth order and the occurrence of any psychiatric disorder.

The comparison of the characteristics of psychiatrically impaired individuals of different birth ranks in several studies does reveal some evidence which indicates that last-born females who are hospitalized for schizophrenia show more flagrant symptomatology and a lower level of social competence, both before and during hospitalization, than do firstborns. Despite the congruence between the apparently disproportionately large number of last-born females among hospitalized schizophrenics and the finding that last-born hospitalized female schizophrenics are sicker and less socially competent before and during hospitalization, the apparent linkage between birth order and schizophrenia among women cannot be taken too seriously. There are several reasons to suspect that the results in question may be due to chance—the rejection of a true null hypothesis. First, the findings have not been replicated in the data of the more recent and more broadly based eight-hospital collaborative study (see Footnote 8). Second, no one who is responsible for the relevant studies, including the present author, has been able to come up with an even moderately gripping explanation of why such a phenomenon should exist.

A similar dearth of reliable results has emerged from the analysis of the prevalence of different birth ranks among different normal populations. The most frequently reported differences between birth ranks—the greater occurrence of firstborns among groups marked by unusual intellectual attainment or occupational achievement—seems to be most parsimoniously explained in terms of differences among social class trends in family size. Furthermore, analysis of the Kohn and Schooler (1969) data that are relevant to occupational mobility shows no significant differences in the level of occupational attainment between firstborn and other birth ranks when the occupational status of the individual is compared with that of his siblings.

Similarly, although many such differences have been hypothesized, the overview of psychological studies comparing the performances of individuals of different birth ranks in various psychological tests and experiments failed to reveal any consistent and replicable pattern of relationship between birth order and normal aspects of personality. Finally, the examination of parents' reports of differences in their treatment of children of different birth ranks, particularly in the data collected by Kohn (1969) and by Kohn and Schooler (1969), failed to reveal any meaningful and replicable pattern of differences.

In 1936 Stagner and Katzoff, noting the extreme paucity of birth order findings in their study of the responses of 430 men on the Bernreuter Personality Inventory, commented:

> The fact that the findings of the present study are largely negative should not be surprising to anyone who has worked with personality problems. The surprising point is that so many presumably thoughtful psychologists have emphasized the importance of birth order in determining personality [p. 345].

More than 30 years later the conclusions of the present article would seem to indicate that Stagner and Katzoff were justified in their doubts about the importance of birth order as a determinant of behavior. However, if they were surprised then at the number of psychologists who took this variable seriously, they would be amazed now. This state of affairs leaves the present writer, who has contributed his share to the intervening literature, with two questions: To what extent does birth order have an effect on any population under any conditions? Why has birth order analysis continued to have such fascination for researchers, and as a corollary to this, are there any reasons why—given the pitifully few replicable findings that have emerged from this relatively large body of research—the study of birth order effects should be continued?

In response to the first of these questions, it does seem possible that birth order has affected personality in other times and other places, particularly where primogeniture has been important. Caudill and Schooler (1969) and Barry (1967) have raised the possibility that where financial and social privilege and the accompanying responsibility of maintaining the family name, property, and integrity are placed on the firstborn, there would be an increase in the likelihood of the firstborn either attaining more or breaking up. Such may well be the case in India,[9] and such may have been the case in Japan in 1958 and possibly even in the United States in the past among traditionally oriented segments of society, but it does not seem to be so in the United States at the present time.

It is also possible that a more consistent pattern of birth order effects would be discovered if other family structure variables were taken into account. Two such potentially relevant variables for which measures have been devised, and some results reported, are density and sex of siblings. A paper by Waldrop and Bell (1964) provides an example of how family density might be measured. It finds a positive relationship between the initiation of contact with others among nursery school children and a measure of family density derived from a factor analysis in which the variables used were the total number of children in the family, the average time span between births, and the time span between the reference child and the next older and next younger sibling.

Several studies have attempted to take into account the effects of the sex of the siblings in different ways. Rosenberg (1965) found an interaction between birth order and sex of siblings. In this interaction, boys in the last half of their sibship who came from families in which they were the only male, or in which the majority of the other siblings were female, had a high self-esteem. However, in the Kohn and Schooler (1969) study such sibling configurations had no significant effects on self-image or on any other variable either by themselves or in interaction with birth order.

Another method of taking into account the sex of siblings is that used by Caudill and Schooler (1969) in their study of birth order effects in Japan. In this study, besides assigning birth order scores in terms of the individual's position in the total sibship, the authors also examined the effect of the individual's birth rank among siblings of the same sex. Thus, for example, an individual could be categorized both as the second of five siblings and the first of three male siblings. In general, the pattern of findings indicated that in Japan at the time of the study same-sex sibship rank was a more potent variable than rank in total sibship.

It is obvious, however, that if the interactions with density of sibship and sex of siblings are fully examined, birth order analysis ceases to be a relatively simple matter. Minimally such an attempt should involve an examination of the effects of differences in the sex and spacing of siblings adjacent to subjects of different birth ranks. When all such combinations are taken into account, the apparently simple independent variable of birth order becomes immensely complex.

Another complication that the present article raises for future birth order studies is the apparent validation it provides for the Price and Hare hypothesis. The apparent validity of that hypothesis requires that future investigators obtain appropriate control groups when attempting to prove the disproportionate prevalence of different birth ranks in different populations. At the very least, it would seem necessary to match controls with the subject population on the bases of age, social class, and family size. Obviously untenable are methods such as the Greenwood and Yule (1914) procedure or the direct comparison of the proportion of firstborn to lastborn individuals, which are both based on the assumption that the probable occurrence of individuals of different birth ranks can be calculated directly from a knowledge of the distribution of birth ranks and family sizes in the sample being considered. Price and Hare's work rules out the possibility of predicting the expected proportions of individuals of different birth ranks within a sample without referring to the effects of trends in the number of families started or average family size among relevant subpopulations.

---

[9] But even here a very recent study by Chakraborty (1970), of 1,426 Indian outpatients, found no birth order differences in schizophrenia, "other psychoses," or psychoneuroses.

Unfortunately, it may well have been the seeming simplicity of birth order as an independent variable that provides the answer to our second question—its attractiveness to researchers. This apparent simplicity and objectivity was probably the main source of the appeal of birth order as an independent variable to those investigators looking for a simple way of controlling genetic effects or an objective way of studying family dynamics. However, if in order to be fruitful the study of birth order necessitates dealing with the complexities of such variables as sex of siblings and family density, or of collecting data on hard-to-define control groups, it loses much of its appeal.

Certainly, unless one is able to make very specific predictions on the basis of an empirically well-grounded theory, the general lack of consistent findings revealed by this review leaves real doubt as to whether the chance of positive results is worth the heavy investment needed to carry out any more definitive studies. On the other hand, I suspect that future investigators, including myself, will not be able to resist the temptation of taking a cheap bet on a long shot by collecting birth order data on their subjects as they pursue studies more central to their interests. And that's how we got into this trouble in the first place (N. B. Schachter, 1959).

## REFERENCES

ALEXANDER, C. N., JR. Ordinal position and sociometric status. *Sociometry,* 1966, **29**, 41–51.

ALTUS, W. D. Birth order and its sequelae. *International Journal of Psychiatry,* 1967, **3**, 23–39.

APPERLY, L. A study of relevant American Rhodes Scholars. *Journal of Heredity,* 1939, **30**, 493–495.

BAKAN, D. The relationship between alcoholism and birth rank. *Quarterly Journal of Studies on Alcohol,* 1949, **10**, 434–440.

BARRY, H., JR. Birth order: Achievement, schizophrenia and culture. *International Journal of Psychiatry,* 1967, **3**, 439–444.

BARRY, H., III, & BARRY, H., JR. Birth order, family size, and schizophrenia. *Archives of General Psychiatry,* 1967, **17**, 435–440.

BAYER, A. E. Birth order and college attendance. *Journal of Marriage and the Family,* 1966, **28**, 480–484.

CAUDILL, W., & SCHOOLER, C. Symptom patterns and background characteristics of Japanese psychiatric patients. In W. Caudill & Tsung-Yi Lin (Eds.), *Mental health research in Asia and the Pacific.* Honolulu: East-West Center Press, 1969.

CHAKRABORTY, A. Birth order and mental illness. *The British Journal of Social Psychiatry,* 1969, **3**, 231–236.

CHEN, E., & COBB, S. Family structure in relation to health and disease. *Journal of Chronic Diseases,* 1960, **12**, 544–567.

CLAUSEN, J. A. Family size and birth order as influences upon socialization and personality. *Bibliography and abstracts.* New York: Social Science Research Council, 1965.

COBB, S. *Emotions and clinical medicine.* New York: Norton, 1950.

DATTA, L. Birth order and early scientific attainment. *Perceptual and Motor Skills,* 1967, **24**, 157–158.

ELLIS, H. *A study of British genius.* London: Hurst & Blackett, 1904.

ERLENMEYER-KIMLING, L., VAN DEN BOSCH, E., & DENHAM, B. The problem of birth order and schizophrenia: A negative conclusion. *British Journal of Psychiatry,* 1969, **115**, 659–678.

EYSENCK, H. J., & COOKSON, D. Personality in primary school-children: 3. Family background. *British Journal of Educational Psychology,* 1969, **40**, 117–131.

FARINA, A., BARRY, H., III, & GARMEZY, N. Birth order of recovered and nonrecovered schizophrenics. *Archives of General Psychiatry,* 1963, **9**, 224–228.

FARLEY, F. H. Birth order, achievement-motivation and academic attainment. *British Journal of Educational Psychology,* 1967, **37**, 256.

GINI, C. Superiority of the eldest. *Journal of Heredity,* 1915, **37**, 37–39.

GLUECK, S., & GLUECK, E. *Unraveling juvenile delinquency.* New York: The Commonwealth Fund, 1950.

GRABILL, W. H., KISER, C. V., & WHELPTON, P. K. *The fertility of American women.* New York: Wiley, 1958.

GRANVILLE-GROSSMAN, K. L. Birth order and schizophrenia. *British Journal of Psychiatry,* 1966, **112**, 1119–1126.

GREENWOOD, M., & YULE, G. A. On the determination of size of family and of the distribution of characters in order of birth. *Journal of the Royal Statistical Society,* 1914, **77**, 179–199.

GREGORY, I. An analysis of familial data on psychiatric patients: Parental age, family size, birth order, and ordinal position. *British Journal of Preventive and Social Medicine,* 1958, **42**, 42.

GREGORY, I. An analysis of family data on 1000 patients admitted to a Canadian mental hospital. *Acta Genetica et Statistica Medica,* 1959, **9**, 54–96.

GREGORY, I. Selected personal and family data on 400 psychiatric inpatients. *American Journal of Psychiatry,* 1962, **119**, 397–403.

GREGORY, I. Alcoholism, family size and ordinal position. *Canadian Psychiatric Association Journal,* 1965, **10**, 134–140.

GUNDLACH, R. H., & RIESS, B. F. Birth order and sex of siblings in a sample of lesbians and nonlesbians. *Psychological Reports,* 1967, **20**, 61–62.

HARE, E. H., & PRICE, J. S. Birth order and family size: Bias caused by changes in birth rate. *British Journal of Psychiatry,* 1969, **115**, 647–657.

Hollingshead, A. B., & Redlich, F. C. *Social class and mental illness: A community study*. New York: Wiley, 1958.

Jones, H. E. The environment and mental development. In L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1954.

Kohn, M. L. *Class and conformity: A study in values*. Homewood, Ill.: Dorsey Press, 1969.

Kohn, M. L., & Schooler, C. Class, occupation, and orientation. *American Sociological Review,* 1969, **34**, 659–678.

Lee, R., Pollin, W., Ladusky, W., & Perlin, S. Prevalence and significance of familial psychopathology in a group of hospitalized schizophrenics. Cited by C. Schooler, Birth order and schizophrenia. *Archives of General Psychiatry,* 1961, **4**, 91–97.

Lorr, M., & Klett, C. J. Constancy of psychotic syndromes in men and women. *Journal of Consulting Psychology,* 1965, **29**, 309–313.

Malzberg, B. *Social and biological aspects of mental disease*. Utica, N. Y.: State Hospital Press, 1940.

Masling, J. Birth order and the need for affiliation. *Psychological Reports,* 1965, **16**, 631–632.

McKeithen, J. E. Patterns of motivation as related to ordinal position in the family. *Dissertation Abstracts,* 1965, **25**, 4845–4846.

Miley, C. H. Birth order research 1963–1967: Bibliography and index. *Journal of Individual Psychology,* 1969, **25**, 64–70.

Moran, G. Notes and comments: Ordinal position and approval motivation. *Journal of Consulting Psychology,* 1967, **31**, 319–320.

Price, J. S., & Hare, E. H. Birth order studies: Some sources of bias. *British Journal of Psychiatry,* 1969, **115**, 633–646.

Rees, A. H., & Palmer, F. H. Factors related to change in mental test performance. *Developmental Psychology Monographs,* 1970, **3** (2, Pt. 2).

Rhine, W. R. Conformity behavior in children as related to socioeconomic status, birth order and level of arousal. *Dissertations Abstracts,* 1966, **27**, 95B.

Roe, A. A psychological study of eminent psychologists and anthropologists and a comparison with biological and physical scientists. *Psychological Monographs,* 1953, **67** (2, Whole No. 352).

Rosenberg, M. *Society and adolescent self-image*. Princeton, N. J.: Princeton University Press, 1965.

Rosenfeld, H. M. Relationships of ordinal position to affiliation and achievement motives: Direction and generality. *Journal of Personality,* 1966, **34**, 467–480.

Sampson, E. E., & Hancock, F. R. An examination of the relationship between ordinal position, personality and conformity: An extension, replication and partial verification. *Journal of Personality and Social Psychology,* 1967, **5**, 398–407.

Schachter, S. *The psychology of affiliation*. Stanford, Calif.: Stanford University Press, 1959.

Schachter, S. Birth order, eminence, and higher education. *American Sociological Review,* 1963, **28**, 757–768.

Schooler, C. Birth order and schizophrenia. *Archives of General Psychiatry,* 1961, **4**, 91–97.

Schooler, C. Birth order and hospitalization for schizophrenia. *Journal of Abnormal and Social Psychology,* 1964, **69**, 574–579.

Schooler, C., & Scarr, S. Affiliation among chronic schizophrenics: Relation to intrapersonal and birth order factors. *Journal of Personality,* 1962, **30**, 178–192.

Schooler, C., & Parkel, D. The overt behavior of chronic schizophrenics and its relationship to their internal state and personal history. *Psychiatry,* 1966, **29**, 67–77.

Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing*. Evanston, Ill.: Row Peterson, 1957.

Skipper, J. K., Jr., & McCaghy, C. H. Strip-teasers: The anatomy and career contingencies of a deviant occupation. *Social Problems,* 1970, **17**, 391–405.

Slater, E. Birth order and maternal age of homosexuals. *Lancet,* 1962, xx, 69–71.

Sletto, R. F. Sibling position and juvenile delinquency. *American Journal of Sociology,* 1934, **39**, 657–669.

Smart, R. G. Alcoholism, birth order, and family size. *Journal of Abnormal and Social Psychology,* 1963, **66**, 17–23. (a)

Smart, R. G. The relationship between birth order and alcoholism among women. *Ontario Psychological Association Quarterly,* 1963, **16**, 9–15. (b)

Solomon, S., & Nuttall, R. Sibling order, premorbid adjustment and remission in schizophrenia. *The Journal of Nervous and Mental Disease,* 1967, **144**, 37–46.

Stagner, R., & Katzoff, E. T. Personality as related to birth order and family size. *Journal of Applied Psychology,* 1936, **20**, 340–346.

Sundararaj, N., & Rao, S. R. Order of birth and schizophrenia. *British Journal of Psychiatry,* 1966, **112**, 1127–1129.

United States Department of Health, Education and Welfare. Public Health Service, National Office of Vital Statistics. Births by age of mother, live-birth order, and color. *Vital statistics—Special reports*. Vol. 5. 1959.

Waldrop, M., & Bell, R. Relation of pre-school dependency behavior to family size and density. *Child Development,* 1964, **35**, 1187–1195.

Warren, J. R. Birth order and social behavior. *Psychological Bulletin,* 1966, **65**, 38–49.

Wolkon, G. H., & Levinger, G. Birth order and need for achievement. *Psychological Reports,* 1965, **16**, 73–74.

(Received November 23, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 3, 176–182

# OCCIPITAL ALPHA REVISITED[1]

THOMAS B. MULHOLLAND[2]

*Perception Laboratory, Veterans Administration Hospital, Bedford, Massachusetts*

The human occipital electroencephalogram is describable quantitatively as an alternation between alpha and little or no alpha. This is the alpha-attenuation cycle. Previous descriptions of the occipital alpha rhythms were based on the assumptions of *singularity* of the response of the electroencephalographic alpha rhythm and *functional equivalence* of all alpha and all intervals of alpha attenuation. Both of these assumptions are critically evaluated against the experimental evidence and a new interpretation is presented. The response of the occipital alpha rhythms is a time series, not a singular response. All "alpha" events are not functionally equivalent, nor are all "no-alpha" events. Their position in the time series must be taken into account.

The history of the study of occipital alpha rhythms in humans can be summed up—as an index of mental processes and behavior, it was overrated; as an orderly, predictable phenomenon, it was underrated. This paper is an attempt to move away from these more extreme positions to one which is more realistic and perhaps more fruitful as well.

The approach is descriptive. A neurophysiological theory of the alpha rhythms is not attempted. Although a descriptive approach is incomplete from a neurophysiological standpoint, it permits a review of the "alpha" phenomenon without preconceptions or "explanations" that are, in fact, unverified. Directions that research must take to resolve some of the problems raised here will be clear to the physiologist. For the psychologist, this paper is a criticism of the assumption that the alpha rhythms are "blocked" or attenuated by visual attention (Mulholland, 1968, 1969a; Storm van Leeuwen, Bickford, Brazier, Cobb, Dondey, Gastaut, Gloor, Henry, Hess, Knott, Kugler, Lairy, Loeb, Magnus, Oller, Daurella, Petsche, Schwab, Walter, & Widen, 1966).

## *The Phenomenon Called Alpha*

The earliest descriptions of the 8–13-cycles-per-second occipital alpha rhythms reflected the way the record looked. Berger (orig. publ. 1930) noticed that alpha appeared in a series of "bursts" alternating with intervals of little or no alpha. This kind of description of the electroencephalogram (EEG) emphasized the salient, obvious features of the recording. Later studies showed that the EEG is in fact a large-scale phenomenon, produced by extensive and functionally diversified brain processes, and that a description consistent with these facts was likely to be most reliable. A general description of the occipital EEG in terms of presence or absence of alpha is consistent with the fact that the EEG is a weighted sum of signals reaching the electrodes from a huge number of cortical cells.

The alpha rhythms are not constant; intervals of little or no alpha occur often. In fact, the alternation between alpha rhythms and little or no alpha is the most reliable feature of the occipital EEG. This alternation is called the "alpha-attenuation" cycle here. There is no mechanism or physiological process implied by the term. It is a name for an observable phenomenon. Despite Berger's early measurement of an alpha-attenuation cycle in normals and pathological cases, it was not pursued further, although Bagchi (1937) noted the effect during steady visual stimulation and termed it "intermittent adaptation."

It is tedious to measure each alpha and no-alpha interval, but this is perhaps not the reason for the neglect of an event-series analysis of the EEG. The series as recorded in the typical clinical situation is extremely variable. In the absence of a method that could reduce the unsystematic, uncontrolled variation, and with a lack of theoretical rationale, it simply did not make sense to measure each interval of alpha or intervals of little-or-no-alpha. Instead, the times alpha was present were summed, usually over a fixed time interval, and the ratio of the alpha-time to total time gave the familiar "percent-time" alpha measure.

---

[1] This article was abstracted from a paper presented at the Colloquium on the Oculomotor System and Brain Functions, Slovak Academy of Sciences, Smolenice, Czechoslovakia, October 19–22, 1970. Sponsored by a grant from the Carnegie Institution of Washington, D. C.

[2] Requests for reprints should be sent to Thomas B. Mulholland, Perception Laboratory, Veterans Administration Hospital, Bedford, Massachusetts 01730.

This kind of statistical treatment assumed that all alpha intervals were, in some essential way, equivalent. For instance, alpha intervals were assumed to be associated with less visual attention compared to intervals of low amplitude, fast activity. The alpha "blocking" responses were also assumed to be functionally equivalent to each other and were pooled together to obtain an average response. The assumption of functional equivalence is considered later in relation to the interpretation of alpha and intervals of low amplitude, fast activity in the EEG, following stimulation.

The response of the occipital EEG to stimulation was also recognized early in the EEG, and the description reflected the way the recorded response looked. It looked as if the response were a singular suppression of alpha following stimulation. These twin assumptions of *functional equivalence,* of all alpha or of all little-or-no-alpha events, and, *singularity* of the response of the occipital EEG to stimulation are not only unsupported by experimental evidence, they have perhaps inhibited the conceptualization of the EEG as a phenomenon alternating between a synchronized and a much less synchronized mode, an alpha-attenuation series.

In our laboratory, "alpha" is defined in terms of (*a*) location of the recording electrodes, (*b*) frequency, (*c*) amplitude, and (*d*) the duration of the rhythms having the proper frequency and amplitude .The electrodes are located on the posterior scalp in occipital and parietal locations, $O_1$, $O_2$, $P_3$, $P_4$, in the standard international nomenclature. The frequency is 8–13 cycles per second. Amplitude is defined with reference to the maximum amplitude of an 8–13 rhythm measured with eyes closed. A rhythm having an amplitude greater than 25% of the maximum "resting" rhythm, and lasting longer than .2 seconds is called "alpha." Everything else is "no-alpha." In the rest of this paper the terms alpha and no-alpha refer to this definition. Measurement of the alpha and no-alpha intervals is to the nearest .1 second. This kind of definition does not imply that the response of the occipital alpha rhythms is an either–or phenomenon. The terms alpha and no-alpha simply mean that a recorded rhythm is consistent with an arbitrary criterion or that it is not. The utility of such a definition is an empirical question. Previous studies using this definition combined with a time series analysis have shown that the response of the occipital alpha rhythms to visual stimulation can be specified reliably and quantitatively.

### *Response to the Parietal-Occipital Electroencephalogram*

A variety of stimulus and ocular variables can modify the alpha-attenuation cycle. Stimulus variables are discussed first.

The response of the alpha-attenuation cycle following a step change in visual stimulation is an initial increase in the average durations of no-alpha ($\overline{\Delta}t_{\text{na}}$), followed by a decrease and an initial decrease of the average alpha durations ($\overline{\Delta}t_{\text{a}}$), followed by an increase (Mulholland & Runnals, 1964a, 1964b). In Figure 1, which is from a study of children, a similar effect can be seen in a steady light condition. With the onset of an increase of visual stimulation, $\Delta t_{\text{na}}$ was increased then decreased to approach an asymptote while $\Delta t_{\text{a}}$ had initially decreased then gradually increased. In these examples and in others we have studied, these empirical results can be generalized and approximated by two functions. A hyperbola describes the average disturbance and recovery of the no-alpha intervals, while a straight line describes the average disturbance and recovery of the alpha intervals:

$$\overline{\Delta}t_{\text{a}} = aN + b$$

$$\overline{\Delta}t_{\text{na}} = \frac{c}{N} + d,$$

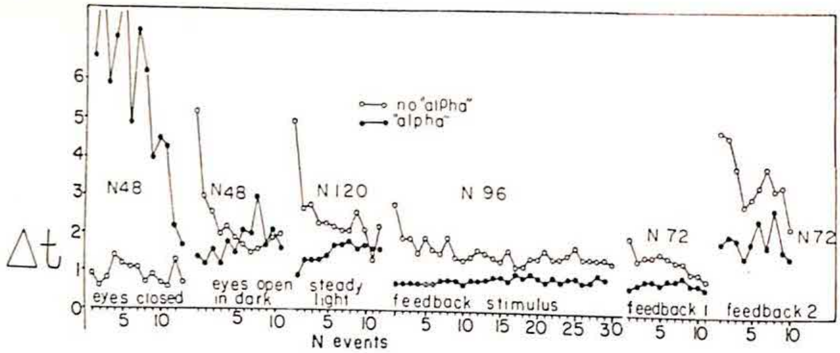


FIG. 1. The response of the average alpha-attenuation cycle to eye opening, step increase, and feedback visual stimulation from 24 children. (For Loop 1, alpha is associated with stimulus ON; no-alpha with stimulus OFF. For Loop 2, no-alpha is associated with stimulus ON; alpha with stimulus OFF.)

where $N$ is the number of alpha and no-alpha events from the onset of stimulation. In our studies $N$ is usually less than 35, and these approximations are for that range of $N$.

The approximately linear and hyperbolic changes in the average alpha, no-alpha durations also occur for disturbances produced by other kinds of stimulation. In our experience all average disturbance and recovery functions can be approximated by these functions, though the particular values of the parameters may change.

### *Feedback Stimulation*

By means of an external path the occurrence of alpha can control the occurrence of a stimulus (Mulholland, 1968; Mulholland & Runnals, 1962a, 1962b). When the stimulus occurrence is contingent upon the alpha, no-alpha durations, each alpha burst can trigger a stimulus pulse and each no-alpha interval can remove the stimulus. Such an arrangement is called Loop 1. In all cases we have studied, the disturbance and recovery of the alpha, no-alpha cycle can be described by the same set of functions presented before. Only the specific values of the average parameters change. Figure 1 illustrates this point. Similar functions have been reported for unpatterned light (Mulholland & Runnals, 1962a, 1962b, 1964a, 1964b), for words flashed on a screen (Mulholland & Davis, 1966), for colored pictures flashed on a screen (Peper, 1970), for television pictures, and for ocular maneuvers (Mulholland, 1969b). In general, the effect of instructional set and variables, such as to "pay attention" to all or certain ones of the stimuli, or to count them, is to change the values of the parameters of the alpha, no-alpha functions (Mulholland & Runnals, 1962b, 1963; Peper, 1970). Also differences between various kinds of stimuli can be described in terms of changes in the specific values of the parameters of the average hyperbolic and linear functions described before.

### *Assumption of Functional Equivalence*

The classical "attention" hypothesis as well as more recent versions assume that all alpha events are equivalent in the sense that they are associated with a state of attention which is always less than that associated with intervals of markedly attenuated alpha or intervals of low amplitude, fast activity. The functions $\bar{\Delta}t_{\mathrm{a}}$ and $\bar{\Delta}t_{\mathrm{na}} = \mathrm{f}\,(N)$ discussed before lead to a critical reappraisal of this assumption. For instance, is the subject likely to be more attentive when $N$ is small and decrease as $N$ gets larger? From this point of view, the $\Delta t_{\mathrm{a}}$, which is the first alpha to return after stimulation, may be associated with a higher level of "attention" than an interval of no-alpha, low amplitude, fast activity occurring as the thirtieth no-alpha event after stimulation. In other words, "attention" may be a function of the linear and hyperbolic functions described before or some combination of these such as $\bar{\Delta}t_{\mathrm{na}}/\bar{\Delta}t_{\mathrm{a}} = \mathrm{f}\,(N)$. There is need for experimental evidence on this question. All alpha may not be associated with the same low level of attention nor all no-alpha (low amplitude, fast activity) intervals with the same high level of attention.

There is another assumption of functional equivalence that refers to within-event processes, the *functional identity* hypothesis. For instance, it is usually assumed for an interval of no-alpha, low amplitude, fast activity that the same level of attention is maintained throughout. Alpha blocking at the beginning is assumed to be associated with the same level of attention as alpha blocking at the end, just before the alpha returns. However, it is possible that just the opposite is true. The level of arousal, activation, or attention in the first 10% of an interval of no-alpha, low amplitude, fast activity may be greater in the first compared to the last 10%, just before alpha returns. In fact, there may be an increase followed by a decrease of cortical activity or "attention" *within* a no-alpha interval. By the careful measurement of behavior during these intervals, or measures of the EEG-evoked potential where the stimulus occurs with controlled delays after the onset of a no-alpha, fast activity interval, it might be possible to estimate and operationally define the time course of the hypothetical "attention" process occurring *within* a no-alpha interval. Similar considerations can be applied to the intervals of alpha (Mulholland, 1968).

### *Ocular and Oculomotor Variables*

The association between alpha rhythm and vision was recognized early in EEG (Adrian, 1943; Adrian & Matthews, 1934) and has been reaffirmed by many others. However, the emphasis has been on "visual attention" and on the so-called arousal and activation pattern of the EEG resulting from complex inhibitory and excitatory changes in unit cell discharges within parietal and occipital cortex associated with processes in the reticular formation of the brain (Jasper, 1958; Moruzzi & Magoun, 1949). These interpretations of the response of the occipital EEG following stimulation emphasized the input, afferent, and associative processes in brain, both specific and nonspecific. It is also likely that efferent, integrative processes occurring in those brain regions from which the parietal-occipital EEG is recorded may also contribute to the response of the occipital-parietal EEG. There is ample evidence, both anatomical and from neurophysiological experiments, to make it a reasonable hypothesis that integrative and efferent processes related to moving and positioning the eyes and to adjusting the lens would be a part of a complex process which is seen in a general, gross way as the disturbance and recovery of the alpha, no-alpha cycle.

The occipital EEG is a gross, general recording from an extensive cortical region; the oculomotor system has extensive representation in Brain Regions 17, 18, and 19, which include the parietal and occipital cortex (Robinson, 1968). Conversely, oculomotor or visual functions which do not involve Regions 17, 18, and 19 in a continuous and extensive process may not have an observable effect on the occipital EEG. For instance, episodic, saccadic movements that involve primarily the frontal eye fields (Robinson, 1968) may not be associated with changes in the recorded occipital EEG alpha rhythm.

The occipital alpha "blocking" response involves processes in the reticular substance of the brainstem tegmentum. Stimulation of this region produces a desynchronized EEG. The oculomotor pathway is located in the same zone of the brainstem tegmentum as the reticular activating system, and stimulation that can cause cortical desynchronization also causes alterations in eye movements (Bender & Shanzer, 1964).

The occipital alpha rhythm is markedly reduced during ocular fixation and pursuit tracking (Adrian & Matthews, 1934; Durup & Fessard, 1935; Mulholland, 1968). These processes involve Regions 17, 18, and 19 (Robinson, 1968), which are the regions from which the parietal-occipital EEG is recorded. For subjects having recordable alpha with eyes open, alpha occurs often when there is no pursuit tracking and the near target appears to be indistinct and blurred following instructions to relax accommodation. When pursuit tracking occurs, even with the target blurred and accommodation relaxed, alpha rhythms are reduced as previous studies show (Dewan & Mulholland, 1969; Mulholland, 1968; Mulholland & Peper, 1971).

Motor functions or sensory-motor integrative functions in other regions of the cortex can cause a desynchronization or reduction of synchronous activity in the EEG record similar to desynchronization of the occipital EEG. Directed movement of the limbs desynchronizes the central m$\mu$ rhythm (Chatrain, Magnus, Petersen, & Lazarte, 1959). Recordings from the precentral gyrus in man show that voluntary movement blocks the precentral beta rhythms (Jasper & Penfield, 1949). These authors stated:

> It would seem that the beta rhythm is characteristic of the activity of the "resting" motor cortex in a manner analogous to the alpha rhythm for the occipital cortex. Activation by voluntary movement seems, therefore, to block the beta rhythm in a manner similar to the well-known effect of visual stimuli upon the occipital alpha rhythm [p. 171].

Here the present author emphasizes the possibility that the analogy drawn by Jasper and Penfield is valid because *oculomotor* processes are involved in the response of the parietal-occipital EEG to visual stimuli.

In a series of experiments we recorded the parietal-occipital EEG while subjects were fixating a stationary target or tracking a moving one with near and relaxed accommodation. The results were consistent with the hypotheses that oculomotor processes originating in Regions 17, 18, and 19 are associated with changes in the alpha, no-alpha cycle. With stationary targets, more alpha occurred during a report of a blurred target combined with accommodative divergence. Much less alpha occurred for reports of a focused target combined with accommodative convergence. For moving targets, least alpha occurred with "focused" tracking; more alpha with "blurred" tracking; and most alpha when the target was reported "blurred" with divergence and no pursuit tracking. Two additional observations are important here: (*a*) some intervals of alpha occurred for some subjects for all conditions; (*b*) saccadic eye movements were not reliably associated with either alpha or no-alpha (Dewan & Mulholland, 1969; Mulholland, 1968; Mulholland & Peper, 1971).

Average as well as individual results can be described in terms of the disturbance and recovery produced following ocular maneuvers. In the experiment just cited, the onset of an ocular maneuver was associated with an instruction from the experimenter. These were "focus–track," "blur–track." "Blur and not track." The series of $\Delta t$ alpha and $\Delta t$ no-alpha events was measured for each condition starting from the initiation of the instruction—ocular maneuver. In general, for the hyperbolic and linear functions described before, $b$ for pursuit tracking was smaller than for conditions of relaxed accommodation and no tracking. For the same conditions, $c$ and $d$ were larger. No definite change occurred for $a$ (Mulholland, 1969b) (see Figure 2).

The lack of an empirical association between saccadic movements and the alpha-attenuation cycle may have been a consequence of the location of the EEG electrodes. Robinson (1968) and others have reported the brain processes for saccades are prominently represented in the *frontal* regions, and we were recording from posterior scalp.

The facilitation of alpha by extreme ocular deviation has been observed in some subjects by Mulholland and Evans (1965, 1966), Mulholland (1969a, 1969b), Fenwick and Walker (1969), and Kris (1968). Chapman, Shelburne, and Bragdon (1970) failed to find evidence for this effect especially with less-than-extreme deviations. Extreme ocular deviation includes variation in a large set of parameters including eye position, lens accommodation, lid closure, and, in some situations, a change of stimulation, although this latter can easily be controlled. Extreme ocular deviation does reliably facilitate occipital alpha for some subjects, and therefore it raises serious questions concerning the set of variables that can modify the alpha rhythm. However, "ocular deviation" cannot be part of a method for answering the question raised because it is a complex set of confounded variables. Ocular deviation, when it is effective in changing alpha, involves changes in accommodation, lack of fixation, and tracking which are consequences of the extreme deviation. However, ocular deviation with fixation and near accommodation may be associated with minimal alpha; deviation with *loss* of fixation and relaxed accommodation may be associated with much more alpha. There is a limit to the utility of EEG—eye-movement studies to resolve these questions. Direct neurophysio-

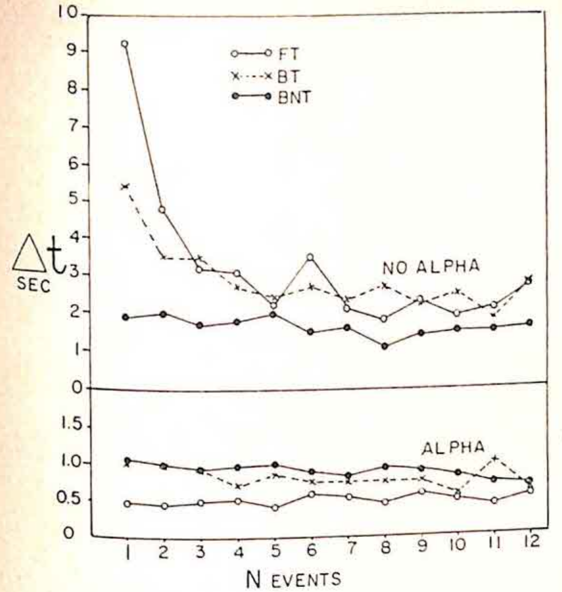


FIG. 2. Disturbance and recovery of the average alpha-attenuation cycle following the start of various ocular maneuvers; moving target (FT = focus-track; BT = blur-track; BNT = blur and not track.)

logical recordings in animals from parietal and occipital cortex and from the effector process in brain, nerve, and eye muscles will perhaps resolve the questions that have been raised. However, it seems clear that an operational definition of visual attention in relation to the occipital EEG refers to the input, feedforward, feedback, and output functions of the visual control systems.

## REFERENCES

ADRIAN, E. D. The dominance of vision. *Transactions of the Ophthalmological Society of the United Kingdom*, 1943, **63**, 194–207.

ADRIAN, E. D., & MATTHEWS, B. H. C. The Berger rhythm: Potential changes from the occipital lobes in man. *Brain*, 1934, **57**, 354–385.

BAGCHI, B. K. The adaptation and variability of response of the human brain rhythm. *Journal of Psychology*, 1937, **3**, 463–485.

BENDER, M. B., & SHANZER, S. Oculomotor pathways defined by electrical stimulation and brain stem lesions in the monkey. In M. B. Bender (Ed.), *The oculomotor system*. New York: Harper & Row, 1964.

BERGER, H. On the electroencephalogram of man. II. In P. Gloor (Trans. & Ed.), Hans Berger on the electroencephalogram of man. The fourteen original reports on the human electroencephalogram. *Electroencephalography and Clinical Neurophysiology Supplement*, 1969 (No. 28), 78. (Orig. publ.: *Journal für Psychologie und Neurologie*, 1930, **40**, 160–179).

CHAPMAN, S., SHELBURNE, S., & BRAGDON, H. EEG alpha activity influenced by visual input and not by eye position. *Electroencephalography and Clinical Neurophysiology*, 1970, **28**, 183–189.

CHATRIAN, G. E., MAGNUS, M. D., PETERSEN, C., & LAZARTE, J. A. The blocking of the rolandic wicket rhythm and some central changes related to movement. *Electroencephalography and Clinical Neurophysiology*, 1959, **11**, 497–510.

DEWAN, E. M., & MULHOLLAND, T. B. The visual control system and the occipital alpha rhythm, an hypothesis. *Electroencephalography and Clinical Neurophysiology*, 1969, **26**, 633. (Abstract)

DURUP, G., & FESSARD, A. L'electrencephalogramme de l'homme. Observations psycho-physiologiques relatives a l'action des stimuli visuels et audifs. *L'Année Psychologique*, 1935, **36**, 1–35.

FENWICK, P. B. C., & WALKER, S. The effect of eye position on the alpha rhythm. In C. R. Evans & T. B. Mulholland (Eds.), *Attention in neurophysiology*. London: Butterworths, 1969.

JASPER, H. H. Recent advances in our understanding of ascending activities of the reticular system. In H. H. Jasper, L. D. Proctor, R. S. Knighton, W. D. Noshay, & R. T. Costello (Eds.), *Reticular formation of the brain*. Boston: Little, Brown, 1958.

JASPER, H. H., & PENFIELD, W. Electrocorticograms in man. Effect of voluntary movement on the electrical activity of the precentral gyrus. *Archiv für Psychiatrie und Zeitschrift Neurologie*, 1949, **183**, 163–174.

KRIS, C. EOG and EEG measurement while learning to position the eyes when the lids are open and closed. *Electroencephalography and Clinical Neurophysiology*, 1968, **24**, 189.

MORUZZI, G., & MAGOUN, H. W. Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, 1949, **1**, 455–486.

MULHOLLAND, T. Feedback electroencephalography. *Activitas Nervosa Superior*, 1968, **10**, 410–438.

MULHOLLAND, T. The concept of attention and the electroencephalographic alpha rhythm. In C. R. Evans & T. B. Mulholland (Eds.), *Attention in neurophysiology*, London: Butterworths, 1969. (a)

MULHOLLAND, T. Occipital alpha activity during accommodative vergence, pursuit tracking and saccadic eye movements. *Electroencephalography and Clinical Neurophysiology*, 1969, **27**, 548. (Abstract) (b)

MULHOLLAND, T., & DAVIS, E. Electroencephalographic activation. Nonspecific habituation by verbal stimuli. *Science*, 1966, **152**, 1104–1106.

MULHOLLAND, T., & EVANS, C. R. An unexpected artefact in the human electroencephalogram concerning the alpha rhythm and the orientation of the eyes. *Nature*, 1965, **207**, 36–37.

MULHOLLAND, T., & EVANS, C. R. Oculomotor function and the alpha-activation cycle. *Nature*, 1966, **211**, 1278–1279.

Mulholland, T., & Peper, E. Occipital alpha and accommodative vergence, pursuit tracking and fast eye movements. *Psychophysiology,* 1971, **8**, 556–575.

Mulholland, T., & Runnals, S. A stimulus-brain feedback system for evaluation of alertness. *Journal of Psychology,* 1962, **54**, 69–83. (a)

Mulholland, T., & Runnals, S. Evaluation of attention and alertness with a stimulus-brain feedback loop. *Electroencephalography and Clinical Neurophysiology,* 1962, **14**, 847–852. (b)

Mulholland, T., & Runnals, S. The effect of voluntarily directed attention on successive cortical activation responses. *Journal of Psychology,* 1963, **55**, 427–436.

Mulholland, T., & Runnals, S. Cortical activation during steady and changing visual stimulation. *Electroencephalography and Clinical Neurophysiology,* 1964, **17**, 371–375. (a)

Mulholland, T., & Runnals, S. Cortical activation by alternate visual and auditory stimuli. *Cortex,* 1964, **1**, 225–232. (b)

Peper, E. Feedback regulation of the alpha electroencephalogram activity through control of internal and external parameters. *Kybernetic,* 1970, **7**, 107–112.

Robinson, D. A. Eye movement control in primates. *Science,* 1968, **161**, 1219–1224.

Storm Van Leeuwen, W., Bickford, R., Brazier, M., Cobb, W. A., Dondey, M., Gastaut, H., Gloor, P., Henry, C. E., Hess, R., Knott, J. P., Kugler, J., Lairy, G. C., Loeb, C., Magnus, O., Oller Daurella, L., Petsche, H., Schwab, R., Walter, W. G., & Widén, L. Proposal for an EEG terminology by the terminology committee of the IFSECN. *Electroencephalography and Clinical Neurophysiology,* 1966, **20**, 293–320.

(Received November 27, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 3, 183–185

# NOTE ON IMMEDIATE MEMORY FOR DIGITS:

## INVARIANCE OVER THE YEARS

HERMAN H. SPITZ[1]

*E. R. Johnstone Training and Research Center, Bordentown, New Jersey*

A review is made of experimental studies of immediate memory for digits in which exposure time per digit was 1 second, plus or minus .3 second, and in which results could be derived as percentages correct at each numerosity level. The survey indicates that the channel capacity (≥90% correct) of unpracticed adolescents and adults of average and above average intelligence is six, plus or minus one. There is some evidence that, at exposure times of 1 second per stimulus array, the span of apprehension falls within this same range.

Because there are few stable facts in psychology's storehouse, Miller's (1956) justly famous article concerning the magical number seven, plus or minus two, aroused great interest. In the course of his article, Miller stated that the span of immediate memory is about seven units in length, and the span of apprehension encompasses about six objects at a glance. These values are certainly reasonable, although Miller cited only a few references to support them. Simon[2] has proposed a magical number of five, rather than Miller's seven. He suggested that higher digit spans for college subjects and for practiced subjects, when compared to grade-school children and unpracticed subjects, are due to greater chunking of the digits by the former groups. In this regard, Martin and Fernberger (1929), on the basis of introspective reports from two college subjects given more than 50 sessions over a 4-month period, concluded that digit span improvement was based on the development of more efficient methods of grouping. Their subjects eventually became proficient in grouping digits by five, but failed when they tried to increase to six the number of digits in each group.

The purpose of the present note is to contribute a more precise value of immediate memory for digits by gleaning from the literature studies conducted under certain narrowly delineated conditions. The span of apprehension is also touched on, but only peripherally.

One of the problems in determining channel capacity (capacity for processing information) involves the use of a criterion. Simon (see Footnote 2) apparently used group means. Miller defined channel capacity as the upper limit at which the observer accurately matches his response to the stimulus. This might be taken to mean 100% accuracy, which would be rather stringent. When a large number of subjects are tested, a few errors by a few subjects should not unduly affect the final determination of channel capacity; nor should a few errors by a single subject over many trials. Consequently, channel capacity is defined here as that level of task difficulty where at least 90% of the subjects, or 90% of a single subject's responses, are completely accurate.

In the typical digit span study, the experimenter presents varying lengths of decimal digits for immediate recall. For present purposes, accuracy level is defined as recall of all digits in the correct order.

A search of the digit span literature was made in order to determine whether there is evidence for a stable channel capacity. Although the history of digit span studies is long and the body of literature is large, results have usually been given in means. In those few studies where results have been given in raw scores or percentages, exposure times of about 1 second per digit were most frequently used. Consequently, this exposure time became a criterion for inclusion of a

---

[1] Requests for reprints should be sent to Herman Spitz, Director of Research, E. R. Johnstone Training and Research Center, Bordentown, New Jersey 08505.

[2] H. A. Simon. How big is a chunk? Measuring information processing parameters that are not directly observable. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, N. J., April 1970.

TABLE 1

Immediate Memory Span[a] for Digits when Exposure Time Is 1 ± .3 Seconds per Digit

| Author and year | Exposure time | Subjects | Presentation method | Response mode | Capacity |
|---|---|---|---|---|---|
| Gates (1916) | .75 sec/digit | 162–164 college | Aud. Seq.<br>Vis. Sim. | Written<br>Written | 6<br>7 |
| Humpstone (1919) | 1 sec/digit | 475 college | Aud. Seq. | Written | 6 |
| Brotemarkle (1924) | | | | | |
| Experiment I | 1 sec/digit | 1263 college | Aud. Seq. | Written | 6 |
| Experiment VI | 1 sec/digit | 261 college | Aud. Seq. | Oral | 7 |
| Guilford & Dallenbach (1925) | .7 sec/digit | 100, probably college | Aud. Seq. | Written | 5 |
| Oberly (1928) | 1 sec/digit | 7 colleagues | Aud. Seq. | Written | 5 |
| Crannell & Parrish (1957) | 1 sec/digit | 24 college | Aud. Seq. | Oral | 5 |
| Dalrymple-Alford (1967) | 1 sec/digit | 10 college | Vis. Seq. | Oral | 5 |
| Spitz, Goettler, & Webreck (1972) | | | | | |
| Experiment III | 1.3 sec/digit | 22 adults | Vis. Sim. | Oral | 7 |
| Experiment IV | 1.3 sec/digit | 44 high school | Vis. Sim. | Oral | 6 |

Note.—Aud. = Auditory; Vis. = Visual; Seq. = Sequential; Sim. = Simultaneous.
[a] ≥ 90% correct.

study in the present survey. In studies that met this criterion, the subjects had little practice on the digit span task; that is, they had not practiced over hours or days. In most, but not all, of the included studies the subjects were individually tested. In some instances other tasks, or variations of the standard digit span task, were given before or after the standard presentation of digits, with no apparent effect on the results of the standard presentation. For the present survey, studies were included in which digits were presented sequentially to the visual modality, sequentially to the auditory modality, or simultaneously to the visual modality, as long as they met the criterion of approximately 1-second per digit exposure time, and as long as results were derivable as percentages correct.

Results of the survey, displayed in Table 1, show a rather encouraging stability over the years. A channel capacity of six, plus or minus one, encompasses all the data, and, incidentally, the magical numbers proposed by Miller (seven) and by Simon (five).

It would be of interest to know whether the channel capacity for span of apprehension (number of like objects correctly perceived at a glance) also falls within this range. On an intuitive basis, one might expect approximately congruent performance on the two tasks when the exposure time for an array of like stimuli is approximately the same as the exposure time per digit. However, most studies have employed shorter than 1-second exposure times. Additionally, few studies have given results in raw scores or percentages correct. Jevons (1871), using himself as a subject in 1,027 trials, threw quantities of beans so that some fell into a box, and when the beans came to rest he estimated their number. Jevons did give his results as percentages, but the exposure time was a "glance" and consequently uncontrolled. Nevertheless, he exhibited 95% accuracy at five beans, and fell to 82% accuracy at six beans.

In studies carried out in our laboratory (Hoats, 1971; Spitz, Hoats, & Holden, 1968) we have used 1-second exposure times and varied the number of dots flashed on a screen. The channel capacity for high school and college students was five to six dots. Consequently, it appears that at a 1-second exposure time six plus or minus one might well encompass the channel capacity for span of apprehension, as well as for immediate memory for digits.[3]

---

[3] Under the same conditions, the channel capacity of educable retardates for digit span and for span of apprehension is four, plus or minus one (Spitz, in press).

## REFERENCES

Brotemarkle, R. A. Some memory span test problems. *Psychological Clinic,* 1924, **15**, 229–258.

Crannell, C. U., & Parrish, J. M. A comparison of immediate memory span for digits, letters and words. *Journal of Psychology,* 1957, **44**, 319–327.

Dalrymple-Alford, E. C. Repetition and immediate memory. *British Journal of Psychology,* 1967, **58**, 63–67.

Gates, A. I. Mnemonic span for visual and auditory digits. *Journal of Experimental Psychology,* 1916, **1**, 393–403.

Guilford, J. P., & Dallenbach, K. M. The determination of memory span by the method of constant stimuli. *American Journal of Psychology,* 1925, **36**, 621–628.

Hoats, D. L. Numerosity discrimination of homogeneous and subgrouped dot patterns by EMR adolescents and normals. *American Journal of Mental Deficiency,* 1971, **76**, 220–224.

Humpstone, H. J. Memory span tests. *Psychological Clinic,* 1919, **12**, 196–200.

Jevons, W. S. The power of numerical discrimination. *Nature,* 1871, **3**, 281–282.

Martin, P. R., & Fernberger, S. W. Improvement in memory span. *American Journal of Psychology,* 1929, **41**, 91–94.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 1956, **63**, 81–97.

Oberly, H. S. A comparison of spans of "attention" and memory. *American Journal of Psychology,* 1928, **40**, 295–302.

Spitz, H. H. The channel capacity of educable mental retardates. In D. K. Routh (Ed.), *The experimental psychology of mental retardation.* Chicago: Aldine, in press.

Spitz, H. H., Goettler, D. R., & Webreck, C. A. Effects of two types of redundancy on visual digit span performance of retardates and varying aged normals. *Developmental Psychology,* 1972, **6**, 92–103.

Spitz, H. H., Hoats, D. L., & Holden, E. A., Jr. Numerosity discrimination of tachistoscopically presented dots by mental retardates and normals. *American Journal of Mental Deficiency,* 1968, **73**, 127–138.

(Received December 18, 1970)

*Psychological Bulletin*
1972, Vol. 78, No. 3, 186–188

# COMMENT ON "EQUIVALENCE OF INFORMATION IN CONCEPT IDENTIFICATION"

LEONARD M. GIAMBRA[1]

*Miami University, Oxford, Ohio*

In this comment D. Arenberg's equivalence-of-information method is considered. First, the equivalence-of-information method is shown to be dependent upon a nonfixed solution procedure, which was first used by J. S. Bruner, J. J. Goodnow, and G. A. Austin in their monograph *A Study of Thinking*. Second, the equivalence-of-information method is shown to be nonapplicable for stimulus universes where there are dimensions that are not binary.

In a recent article Arenberg (1970) described a method for maintaining "equivalent amounts of information in concept identification when subjects select instances [Arenberg, 1970, p. 355]." In this comment I would like to make the following points relative to Arenberg's article: (*a*) Arenberg's method is dependent upon a nonfixed solution procedure, which was first used by Bruner, Goodnow, and Austin (1956). (*b*) The outcome of the method, which results in informational equivalence for stimulus selections with different concept rules, is applicable only to stimulus universes of binary-valued dimensions.

## *Origin of the Arenberg Method*

The key to Arenberg's general method is his refusal to fix the relevant dimensions of the unknown concept in any problem. By doing this, logically equivalent stimulus selections (these are stimulus instances that have the *potential* for providing equal amounts of information to the subject) can all be given the same informational content. Equivalence is achieved by the experimenter by his being able to quasi-arbitrarily designate as exemplar or nonexemplar each selection instance. This nonfixed solution procedure was first used by Bruner et al. (1956) in their outstanding monograph *A Study of Thinking* (see pages 119–121). Bruner et al.'s purpose was to present their subjects with a series of concept identification problems in which the nature of the stimulus instances selected was predominately exemplar or predominately nonexemplar. What Arenberg has done is to take this procedure and its goal of controlling the exemplar status of selected stimulus instances and apply his own criteria for designating each instance as an exemplar or nonexemplar—his criteria being, of course, achievement of equivalence of information.

## *Limitation on the Generality of the Arenberg Method*

For illustrative purposes Arenberg confined himself to a stimulus universe comprised of instances completely describable by four binary-valued dimensions. Arenberg also confined himself to conjunction and inclusive disjunction concept rules. Each rule had two relevant dimensions, and the subject knew the rule being used but was to determine the appropriate relevant dimensions used with the rule (commonly called the attribute identification task). With a four binary-valued dimensional universe there are 24 possible pairs of relevant dimensions that may be used with either the conjunction or the inclusive disjunction rule. Arenberg demonstrated that the number of possible pairs of relevant dimensions is reduced to 18 by designating the first stimulus selection a nonexemplar when the rule is conjunction and an exemplar when the rule is inclusive disjunction. Thus equivalence of information is achieved for identical selections under these two concept rules. The number of remaining possibilities is reduced to six pairs if the exemplar and nonexemplar status's are reversed. Arenberg further demonstrated that this equivalence of information can be maintained for all selections under these two different concept rules simply by never designating the selections as both exemplars or nonexemplars. Thus the basic principle for maintaining an equivalence of information for identical stimulus selections with the two concept rules is that the number of possible pairs of relevant dimensions eliminated when the rule is conjunction must equal the number eliminated when the rule is inclusive disjunction when the exemplar status's of the stimulus instances selected are different. This principle is demonstrated here as operative only in a stimulus universe of binary-valued dimensions.

---

[1] Requests for reprints should be sent to Leonard M. Giambra, Department of Psychology, Miami University, Oxford, Ohio 45056.

First let us consider a stimulus universe comprised of instances describable along three dimensions and where each dimension has three values. If the number of dimensions relevant to a rule is two, then the number of possible pairs of relevant dimensions is $3\ (3+3)+3\ (3)=27$. If the rule is conjunction and the first instance selected is designated as a nonexemplar, then the number of possible pairs of relevant dimensions (i.e., concepts) eliminated is 3, the number remaining is 24. Had the instance been designated as an exemplar, then the number of concepts eliminated would have been 24 and only 3 would have remained. If the rule is inclusive disjunction and the first instance selected is designated as an exemplar, then the number of possible concepts eliminated is 12, the number remaining is 15. Had the instance been designated as a nonexemplar, then the number of concepts eliminated would be 15 and the number remaining 12. Thus it is *not* possible to manipulate the exemplar or nonexemplar status of the selected stimulus instance so that the number of concepts remaining (or eliminated) with the conjunction rule can be made equal to the number remaining (or eliminated) when the rule is inclusive disjunction. Hence the equivalence-of-information method, does not result in equating information processed using the conjunction or inclusive disjunction rule with ternary-valued dimensions. The following generalizes this conclusion to any arbitrary (but complete) universe with at least one nonbinary dimension in it.

Let $X_i$ $(i=1, \ldots, L)$ represent the number of values associated with the $i$th dimension of an $L$-dimensional stimulus universe. If $L$ is three then the number of possible conjunction or inclusive disjunction concepts in this universe is $X_1\ (X_2+X_3)+X_2\ (X_3)$. If $L$ is four then the number is $X_1\ (X_2+X_3+X_4)+X_2\ (X_3+X_4)+X_3\ (X_4)$ or equivalently:

$$X_1 \sum_{i=2}^{4} X_i + X_2 \sum_{i=3}^{4} X_i + X_3 \sum_{i=4}^{4} X_i = \sum_{j=1}^{3} X_j \sum_{i=j+1}^{4} X_i.$$

The general formula for the number of possible conjunctions or inclusive disjunctions is:

$$\sum_{j=1}^{L-1} X_j \sum_{i=j+1}^{L} X_i. \qquad [1]$$

Next consider that the first stimulus instance has been selected. If the concept rule is conjunction and the stimulus instance is designated as a nonexemplar, then the number of concepts *eliminated is*:

$$\frac{L!}{2!\,(L-2)!}. \qquad [2]$$

The number of concepts *remaining* would be Formula 1 − Formula 2. Had the stimulus instance been designated as an exemplar, then the number remaining would be determined by Formula 2 and the number eliminated by Formula 1 − Formula 2.

If the concept rule is inclusive disjunction and the stimulus instance is designated as an exemplar, then the number of concepts *eliminated* is:

$$\sum_{j=1}^{L-1} (X_j-1) \sum_{i=j+1}^{L} (X_i-1). \qquad [3]$$

The number of concepts *remaining* would be Formula 1 − Formula 3. Had the stimulus instance been designated as a nonexemplar, then the number remaining would be determined by Formula 3 and the number eliminated by Formula 1 − Formula 3.

As previously shown, the equivalence-of-information method only performs its task (for the two concept rule case) when the numbers of concepts remaining (or eliminated) by a stimulus instance selection can be equated by designating the instance for one concept an exemplar (or nonexemplar) and for the other concept a nonexemplar (or exemplar). That is, the equivalence-of-information method will work when only Formula 1 = Formula 3. One can easily demonstrate that Formula 1 can equal Formula 3 only when there are two values per dimension.[2] That is, Formula 1 = Formula 3 if and only if $X_i$ is two for *all* values of $i$. Thus the conclusion must be that the equivalence-of-information method is not applicable for stimulus universes with any dimensions having three or more values to it. The reader can also in a like manner demonstrate to himself that this statement would be true for conjunction and inclusive disjunction concepts involving three relevant dimensions and $L$ greater than three. The statement would also be true for concepts of any number of relevant dimensions (and $L$ greater than this value).[3] Finally, this statement would also remain true with most other concept rules as the reader can easily demonstrate by arbitrarily selecting any two rules, regardless of number of relevant dimensions.

---

[2] Strictly, this is true only for all values of $L$ greater than two. The case where the number of dimensions of the universe is two results in Formula 1 = Formula 3 for any value of $X_i$ since with two dimensions the number of concepts eliminated on all trials is always one or all but one, and this is independent of the number of values per dimension. One may properly consider the case when $L$ is two as a trivial limiting one.

[3] This may be done by generalizing Formulas 1, 2, and 3.

## REFERENCES

Arenberg, D. Equivalence of information in concept identification. *Psychological Bulletin*, 1970, **74**, 355–361.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking*. New York: Wiley, 1956.

(Received January 19, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 3, 189–208

# STIMULUS COMPOUNDING IN FREE-OPERANT AND CLASSICAL CONDITIONING

## A REVIEW AND ANALYSIS[1]

STANLEY J. WEISS[2]

*The American University*

Free-operant and classical conditioning studies reporting additive summation, suppressive summation, and response averaging to compounded stimuli are reviewed, and a stimulus control model applicable to additive and suppressive summation in both paradigms is presented. A symmetrical composite-stimulus continuum, defined by the on–off states of the discriminative or conditioned stimuli controlling behavior in training, is seen to be common to both types of summation, with the functions of the all-on and all-off continuum extremes interchangeable. Moreover, additive summation, suppressive summation, and response averaging appear to be dependent on discrimination training, and determined by the history of, as well as the immediate relations between, response outputs and reinforcement densities conditioned to particular stimulus values along the composite training dimension. This evidence, and the stimulus control model derived therefrom, are used to functionally relate summation to generalization peak shift and to suggest that these additive, suppressive, and averaging effects to compounded stimuli might further help in clarifying, and in placing in contextual perspective, what are currently thought of as "excitatory" and "inhibitory" mechanisms in learning.

This review presents the current status of stimulus-compounding research in free-operant and classical conditioning,[3] and through this presentation it attempts to develop a comprehensive functional analysis of the variables involved in determining the various behavioral resultants reported within this experimental paradigm. The scope of the present paper does not extend to multiple-response or multiple-choice procedures, exemplified by the experiments of Schoeffler (1954) and Atkinson, Calfee, Sommer, Jeffrey, and Shoemaker (1964), but instead focuses on unitary response paradigms where variations in rates or magnitudes of response are of primary concern. Ultimately, the stimulus control analysis formulated is (*a*) discussed in terms of some current conceptions of "excitation" and "inhibition" and (*b*) used to relate stimulus compounding research to other problems under investigation in the area of stimulus control.

In stimulus-compounding research within free-operant and classical conditioning, subjects are customarily brought under the *independent* control of two or more stimuli, usually from different modalities; then, in testing, the response forthcoming to the simultaneous presentation of two or more of the independently conditioned stimuli is compared to the responses controlled by single-stimulus presentations. This procedure should be clearly distinguished from compound-stimulus conditioning, recently reviewed by Razran (1965) and Baker (1968). In compound-stimulus conditioning subjects are conditioned to a multielement stimulus, and the dependent measure is customarily the behavior controlled by the respective elements. This procedure is the mirror image of stimulus compounding.

In the two-stimulus case there are four possible outcomes to stimulus compounding. These are additive summation, averaging, suppressive summation, and no difference in the response to individually presented and compounded stimuli. For additive summation to be evidenced, there is, at the minimum, at least a greater response forthcoming to the compound stimulus than to either individually presented component. When averaging occurs, the response to the compound is somewhere intermediate between those controlled by the two independently conditioned stimuli. Suppressive summation is observed when a lesser response is controlled by the compound than by either of the stimuli presented singly. These several behavioral resultants to stimulus compounding are discussed in turn and compared.

---

[1] The preparation of this manuscript was supported in part by Grant MH-16853 from the National Institute of Mental Health, United States Public Health Service.

[2] Requests for reprints should be sent to Stanley J. Weiss, Department of Psychology, American University, Washington, D. C. 20016.

[3] Due to the limited availability of the Russian literature in English, stimulus compounding within classical conditioning is, unfortunately, probably not covered as extensively as the free-operant research in that area. Nevertheless, the classical conditioning studies presented are consistent enough in their findings, with exteroceptive conditioned stimuli, to adequately support the analyses presented.

## CLASSICAL AND FREE-OPERANT CONDITIONING EXPERIMENTS DEMONSTRATING ADDITIVE SUMMATION

### *Classical Conditioning*

The first studies reporting additive summation were done in Pavlov's laboratory in the early twentieth century. Kimble (1961) cited a classical conditioning study performed by Pavlov in which one conditioned stimulus, the odor of oil of camphor, ordinarily evoked 60 drops of saliva; the other stimulus, a mild shock, usually produced 30 drops. When these stimuli were presented together they evoked 90 drops.

Razran (1939) in his survey of studies in configural conditioning referred to two other Russian experiments that demonstrated additive summation. Leporsky[4] found that simultaneous presentations of several stimuli, each of which had been previously conditioned separately to the same response, produced conditioned responses of greater magnitude than any application of a single stimulus or of a combination of the same but fewer stimuli. Razran reported that Leporsky's results were corroborated by Yakovleva (1927) who also noted that the augmentation by summation is gradually diminished if the combination is continually reinforced. A similar decrease in response enhancement to compound-stimulus presentations was reported by Wolf (1963) in a free-operant paradigm where responding was maintained by food reinforcement.

Hull (1940) detailed a classical conditioning experiment employing the galvanic skin response (GSR) in which eight human subjects were first conditioned with shock to respond to weak light, and next to a weak vibratory stimulus that was paired with the shock. Although the mean reaction, in millimeters, was approximately 3.55 to the light or vibrator separately, the compound stimulus elicited a GSR of 4.40 millimeters. Konorski (1948, pp. 112–113) also reported response enhancement upon the joint presentation of any two of three conditioned stimuli independently associated with food—with the relative enhancement greater in those situations where, through reduced deprivation and/or decreased stimulus intensity, the separately presented conditioned stimuli controlled minimal salivation.

Grings and O'Donnell (1956) differentially conditioned the GSR of 32 human subjects to three classes of colored-light stimuli. These stimuli were reinforced consistently, not paired with reinforcement, or not presented during training. During the test period, response magnitude was observed to pairs of stimuli. A pair of previously reinforced stimuli produced a significantly greater response than a reinforced stimulus paired with a nonreinforced or a nonpresented stimulus. The difference between the last two paired conditions was questionable. Grings, Uno, and Fiebiger (1965) directly replicated the Grings and O'Donnell (1956) experiment and confirmed that on the first test trial response magnitude tended to be ordered according to the principles of response summation. Unfortunately, since their conditioned stimuli (CS+) were continuously reinforced during training, the response elicitation capacity of those stimuli rapidly extinguished during testing.

Timofeev (1959) mentioned results of Vediaev who reported increased conditioned responses (CRs) in both fishes and hens to the simultaneous presentation of independently conditioned exteroceptive stimuli. When Timofeev (1959) compounded independently, and sequentially, conditioned exteroceptive (40-watt light) and interoceptive (stretching of the wall of the anterior part of the alimentary tract via rubber tube insert) stimuli, additive summation was reported in fishes, but not in hens. Timofeev went on to cite that data similar to and even more pronounced than those forthcoming in his experiment with birds were reported in 1952 by Airapet'iants. When Airapet'iants presented an exteroceptive and an interoceptive agent simultaneously to dogs, conditioned as well as unconditioned reactions were inhibited. The difficulties in comparing methodologically and technically exteroceptive-exteroceptive and exteroceptive-interoceptive control, and the scarcity of detailed empirical information, forbids comparison and analysis by the author of these two types of control. For experimental purposes this review is primarily concerned with exteroceptive stimulus control in the training paradigms of both operant and classical conditioning studies exploring stimulus compounding.

---

[4] N. I. Leporsky. Materials to the physiology of conditioned inhibition. Unpublished thesis, St. Petersburg, Russia, 1911.

### *Free-Operant Conditioning*

*Variable-interval schedules.* Weiss (1964; 1971, Experiment 2) and Wolf (1963, Experiment 1) demonstrated within a multimodal training situation that additive summation could be produced in free-operant as well as classical conditioning. They trained rats to respond on a variable-interval (VI) food reinforcement schedule when a tone or a light was present. (On a VI schedule a reinforcement is produced by the first response occurring some average time, quantitatively defining the particular VI schedule, from the previous reinforcement.) During the simultaneous absence of these stimuli, light-out no-tone, extinction was in effect. In testing, all subjects summated to the simultaneous presentation of tone and light; that is, they emitted more responses to this test condition than to light or tone individually presented.[5] Weiss (1964) also reported summation when generalized stimuli were compounded.

*Fixed-interval schedules.* Miller and Ackley (1970) and Miller (1971) produced summation when they compounded $S^D$s individually associated with fixed-interval (FI) schedules. (On an FI schedule, a reinforcement is available for the first response emitted a specified time after the previous reinforcement.) Miller and Ackley trained four rats on a multiple FI FI extinction schedule where FI periods were signaled by tone or light, and $S^{\Delta}$ by the simultaneous absence of tone and light. In spite of substantial rate overlap between single- and compound-stimulus presentations evidenced in test records, the overall test data reveal that (*a*) compound-stimulus presentations controlled a summative effect over the entire fixed-interval period for the three interval schedules these authors employed; and (*b*) for total test data, this summative effect could be adequately described as a simple additive function of the rates controlled by the single FI $S^D$s.[6]

*Avoidance contingencies.* Miller (1969b) separately conditioned a light and a buzzer to control a shock avoidance response in a shuttle box. Those rats that demonstrated a high percentage of avoidance responses, and fairly short latencies, to the single stimuli exhibited significantly shorter avoidance latencies to these stimuli compounded.

Riess (1969) trained rats on a Sidman schedule in a shuttle box where hurdle crossing was the avoidance response. (On a Sidman schedule shocks are unsignaled, but a response postpones shock for a prescribed time—in this instance, 20 seconds.) Subsequently, with the avoidance response restricted, subjects were classically fear conditioned. Tone and light *independently* signaled unavoidable shock.

---

[5] Cornell and Strub (1965), Miller (1971), and Wolf (1963, Experiment 2) also demonstrated summative effects to compounded stimuli when the independently conditioned discriminative stimuli were differentially located lights—with Miller (1971) reporting that an intermodal compound maintained a higher percentage of combined component stimulus response frequencies than did an intramodal compound. Intensity dynamism effects per se have been conclusively shown not to be a factor in multimodal summation (Weiss, 1969a), and Miller (1971, Experiment 1) recently attempted a systematic investigation of this variable's effect during stimulus compounding when the compound stimulus turns out to be a greater intensity of physical energy to a single modality. He interpreted his results as minimizing the importance of compound intensity effects in intramodal compounding, although his between-intensity groups comparison was confounded by the training and testing experienced by the high-intensity group in a previous experiment. However, test performance of low- and moderate-intensity groups, who started the experiment naive, does support Miller's conclusion.

[6] Weiss, at the 1968 Convention of the American Psychological Association in San Francisco, reported a similar additive function when he compounded stimuli controlling responding on differential-reinforcement-of-low-rate (DRL) schedules. (A response is reinforced on a DRL schedule if a minimum specified time has elapsed between responses.) Five animals were trained on a multiple DRL 18-second DRL 25-second extinction schedule with tone and light counterbalanced over DRL schedules, and $S^{\Delta}$ signaled by the simultaneous absence of tone and light. Also in the positive reinforcement situation, Miller and Price (1971) showed that when two stimuli each maintain a certain speed of running, their combination controls a significantly greater speed.

Thereafter, when tone or light were superimposed on the Sidman base rate, conditioned acceleration, an increased avoidance rate, was observed (cf. Sidman, Herrnstein, & Conrad, 1957). However, conditioned acceleration to compounded tone and light was reliably higher—in 22 of 24 daily subject performances—than acceleration to the individually presented stimuli. Riess thus demonstrated summation of conditioned acceleration.

Emurian and Weiss (1972) reported additive summation when they compounded $S^D$s signaling a Sidman avoidance contingency during an extinction test as well as over a substantial majority of 28 maintenance test sessions in which the avoidance contingency was effective during single- and compound-stimulus presentations. Most striking was the fact that subjects received fewest shocks in the compound on 85% of the maintenance tests. In training, a tone and a light each independently signaled that a Sidman avoidance schedule was in effect and maintained responding, while light-out no-tone was a shock-free period that controlled response cessation. Discrimination and shock avoidance criteria were stringent. The relations between test response outputs to single (tone or light) and compound (tone plus light) test presentations appeared generally similar to those reported by Weiss (1969a) when VI $S^D$s were compounded. This led Emurian and Weiss (1970) to conclude that free-operant additive summation is not incentive-specific.

In the Riess (1969) and Emurian and Weiss (1972) studies, rate of avoidance responding was substantially higher during tone or light than during the simultaneous absence of these stimuli. Emurian and Weiss (1972) pointed out that the same general rate relations exist in training between these respective stimulus conditions when additive summation was reported to tone plus light when responding to tone and light was maintained by a food rather than a shock incentive. Emurian and Weiss (1972) therefore concluded that the discrimination training leading to those *differential rates* might be of prime importance in the production of additive summation to compounded stimuli. If this is the case, an interpretation of summative avoidance in terms of an interaction of "anxiety states" may not be warranted, or at best would be incomplete.

## CONSIDERATION OF STIMULUS CONTROL AND DISCRIMINATION TRAINING IN ADDITIVE SUMMATION PARADIGMS

Weiss (1969a) has pointed out that although additive summation has been empirically established in both free-operant and classical conditioning, its dynamics remained obscure. To clarify them, he analyzed the stimulus relations between the various experimental conditions of training and testing in both classical and free-operant summation paradigms and discovered some startling consistencies.

### *Composite-Stimulus Control Analysis*

Similar differential training is common to both classical and free-operant experiments demonstrating additive summation. In classical conditioning studies the presentation of an unconditioned stimulus (UCS) was signaled by the independent presentation of $CS_1$ or $CS_2$ and the simultaneous absence of these stimuli $(\overline{CS}_1 + \overline{CS}_2)$ signaled extinction. That extinction was occurring to this latter condition was first reported by Pavlov who wrote:

> when conditioned reflexes are being established in dogs for the first time, it is found that the whole experimental environment . . . acquires at first conditioned properties. . . . But later on, when the special reflex to the single, definite and constant stimulus has appeared, all the other elements of the environment gradually lose their significance most probably on account of a gradual development of internal inhibition [1927, p. 115].

In the instrumental studies described above, reinforcement periods in training were signaled by the independent presentation of $S^D{}_1$ or $S^D{}_2$. The simultaneous absence of these stimuli $(\overline{S^D}_1 + \overline{S^D}_2)$ customarily was systematically nonreinforced and therefore came to control response cessation just as $(\overline{CS}_1 + \overline{CS}_2)$ in the classical conditioning studies ceased to elicit conditioned responses. Furthermore, a composite comparison between the stimuli that control summation in these studies and those that are associated with reinforcement and extinction reveals additional consistencies.

Weiss (1969a) noted:

> [A] composite continuum is defined by the on-off states of all relevant discriminative or conditioned stimuli, one

TABLE 1

Schematic Representation of Classical and Free-Operant Multimodal Additive Summation Experiments

| Composite stimulus continuum | Light-out No-tone $(\bar{T} + \bar{L})$ | Tone Light-out $(T + \bar{L})$ | or | Light No-tone $(L + \bar{T})$ | Tone and light $(T + L)$ |
|---|---|---|---|---|---|
| 1. Extinction signaled by all-off $(\bar{T} + \bar{L})$ extreme | | Training | | | |
| (*a*) Classical conditioning experiments[a] | Extinction | $CS_1$ | | $CS_2$ | Summation obtained |
| | | | | Test | |
| | | Training | | | |
| (*b*) Free-operant experiments[b] | Extinction ($S^\Delta$) | $S^D{}_1$ | | $S^D{}_2$ | Summation obtained |
| | | | | Test | |
| 2. Extinction signaled by all-on $(T + L)$ extreme | | No data available | | | |
| (*a*) Classical conditioning experiments | | | | | |
| | | | | Training | |
| (*b*) Free-operant experiments[c] | Summation obtained | $S^D{}_1$ | | $S^D{}_2$ | Extinction ($S^\Delta$) |
| | | Test | | | |

[a] Leporsky (1911), Yakovleva (1921), Pavlov (in Kimble, 1961), Hull (1940), Konorski (1948), Timofeev (1959).

[b] Wolf (1963, Experiment 1), Weiss (1964; 1971, Experiment 2); Emurian & Weiss (1972), Miller & Ackley (1970), Miller (1971).

[c] Weiss (1969a), Meltzer & Freeman (1970).

> continuum extreme anchored to the all-off state, the other to all-on. The points between these continuum extremes are scaled with reference to the number of elements common to each . . . , the number of points on any composite scale being equal to the number of independently conditioned $S^D$s, or CSs, plus one [pp. 22–23].

Using tone (T) and light (L) as *representative* stimuli to make the example concrete, one notes that in both the reported free-operant and classical conditioning studies yielding additive summation the composite continuum starts at the all-off extreme signaling extinction $(\bar{T} + \bar{L})$, progresses through the individual presentation of each $S^D$ or CS that maintains responding $(T + \bar{L}$ or $L + \bar{T})$, to the all-on extreme $(T + L)$ that yields summation. A representative two-element, tone and light, three-step composite continuum is shown in the top row of Table 1 which *schematically* represents the training and test conditions of those multimodal classical and free-operant summation experiments that associated the absence of physical stimulus energy with extinction.[7] Comparable values along this dimension are associated with similar experimental conditions in both paradigms.

Weiss (1969a) pointed out that a binary analysis of the on–off stimulus relations between reinforcement-associated stimuli and extinction in the top row of Table 1 reveals that each $S^D$ or CS contains an $S^\Delta$ (extinction) element. But since additive summation was clearly demonstrated to $T + L$, or its equivalent, in testing, it appeared that the components of the $T + \bar{L}$ and $L + \bar{T}$ reinforcement-associated stimuli controlling responding were those $S^D$ or CS elements not also coincident with extinction, T and L, respectively. Thus, Weiss (1969a) observed:

---

[7] Although multimodal composite dimensions are being primarily discussed because they (*a*) are most prominent in the stimulus-compounding literature and (*b*) reduce the probability of interference or generalization between independently conditioned stimuli, this analysis can also be applied to experiments reported by Wolf (1963, Experiment 2) and Cornell and Strub (1965) who used different lights, $L_1$ and $L_2$, as their $S^D$s. The composite continuum here would be: $(\bar{L}_1 + \bar{L}_2)$, $(L_1 + \bar{L}_2)$ or $(\bar{L}_1 + L_2)$, $(L_1 + L_2)$. This same reasoning might be applied conceptually to the unimodal classical conditioning studies of Grings and O'Donnell (1956) and Grings et al. (1965). Unfortunately, confirmatory data are not available. However, it should be emphasized that for an unambiguous application of the composite-stimulus control model in stimulus compounding (*a*) the characteristics of the discriminative or conditioned stimuli controlling the various responses in training must be clearly evident and not open to question or interpretation; and (*b*) the independently conditioned stimuli when presented simultaneously must not coalesce so as to camouflage the identity of the underlying composite elements. That is why multimodal designs are preferred in compounding research.

> summation was forthcoming to the composite stimulus composed entirely of elements unique to an $S^\Delta$ [or CS] function, which, lacking discriminated $S^\Delta$ elements, was further removed from $S^\Delta$ on the postulated continuum than either of the other test stimuli [p. 22].

Weiss' (1969a) position might be restated so as to make his underlying assumptions somewhat more explicit. When discrimination training occurs we have a situation that can be described as follows: (*a*) $\bar{T} + \bar{L}$, associated with extinction, controls response cessation (behavior$_1$); (*b*) During $T + \bar{L}$, responding, maintained by reinforcement, consists of response under control of T (behavior$_2$) and response cessation (behavior$_1$) under control of L. Therefore, rate for this condition is lower than would be the case if $\bar{L}$ was not in effect. (*c*) Similarly, during $L + \bar{T}$ responding, also maintained by reinforcement, consists of response under control of L (behavior$_3$) and response cessation (behavior$_1$) under the control of $\bar{T}$. Therefore, response rate in this condition is lower than would be the case if $\bar{T}$ were not in effect. (*d*) Finally, during the $T + L$ test condition, response rate is higher (additive summation) because responding consists of a mixture of response under the control of T (behavior$_2$) and response under the control of L (behavior$_3$) which yields a higher rate than the mixture of either behavior$_2$ or behavior$_3$ with "response cessation" (behavior$_1$). Of course according to the same reasoning if $\bar{T} + \bar{L}$ controls behavior that results in *higher* rates of responding than $T + \bar{L}$ or $L + \bar{T}$, then one would expect that when $T + L$ was presented, alternation of responses to T and to L would result in lower rates than to either $T + \bar{L}$ or $L + \bar{T}$. This latter prediction is confronted in subsequent sections of this paper (see "Suppressive Summation" and "Rate Relations during Additive and Suppressive Summation"). A prediction from this model more directly concerned with additive summation is what next concerned Weiss (1969a).

From the foregoing analysis Weiss (1969a) then argued that if the stimulus control relations postulated for the two-element composite dimension did exist, the continuum's behavioral symmetry could be tested through a free-operant experimental design that reversed the locus of $S^\Delta$ to the $T + L$ continuum extrem while maintaining $T + \bar{L}$ and $L + \bar{T}$ as the independently reinforced $S^D$s. Now the elements of each $S^D$ controlling responding should be $\bar{L}$ and $\bar{T}$, respectively. This derives from a mirror-image analysis of those attentional processes postulated when $S^\Delta$ was $\bar{T} + \bar{L}$. Now that composite condition composed entirely of elements unique to an $S^D$ function, thereby lacking $S^\Delta$ elements and thus furthest removed from $S^\Delta$ on the composite scale, is $\bar{T} + \bar{L}$, or "stimulus absence." Weiss (1969a) therefore predicted that additive summation should be forthcoming to the $\bar{T} + \bar{L}$ test condition. The design of this experiment is shown in row 2b of Table 1.

Weiss (1969a) did get summation to $\bar{T} + \bar{L}$ to a degree that was directly comparable to that forthcoming to $T + L$ in a mirror-image free-operant summation study that employed similar training and testing techniques. He therefore concluded:

> The postulation of a symmetrical composite stimulus dimension in summation designs has been supported. [Additive] summation will occur to the stimulus complex representing either extreme of a two-modality composite continuum when "single" stimulus presentations are reinforcement-correlated if the opposite extreme is conditioned as $S^\Delta$. Therefore the continuum is symmetrical, with the functions of the two extremes interchangeable [Weiss, 1969a, p. 25].

Extrapolating from Weiss' (1969a) composite-stimulus control model and these results, it becomes evident that a "composite extreme" is defined in relation to the behavior it controls and not absolutely. It is not logically necessary for $\bar{T} + \bar{L}$ and $T + L$ to be dimension extremes. $T + \bar{L}$ and $L + \bar{T}$ could serve equally well if *either* were associated with extinction and controlled response cessation while $\bar{T} + \bar{L}$ *and* $T + L$ maintained responding through reinforcement. In the former case, where $T + \bar{L}$ was $S^\Delta$, "the composite stimulus composed entirely of elements unique to an $S^D$ function" is $L + \bar{T}$, and it should control additive summation in testing. On the other hand, when $S^{\Delta}$ is $L + \bar{T}$, while $\bar{T} + \bar{L}$ and $T + L$ remain the $S^D$s maintaining responding, it is $T + \bar{L}$ that should control additive summation since it contains only composite "elements unique to an $S^D$ function." Experiments substantiating hese latter predictions have not yet been ported in the literature.[8]

### *Influence of $S^{\Delta}$ Training*

In addition to establishing a stimulus control model within which the additive summation studies of Table 1, as well as other suppressive summation designs to be discussed later, neatly fit, Weiss' (1969a) findings also implicitly focus upon the importance of the $S^{\Delta}$ function in the discrimination training of these additive summation studies. Discrimination training, customarily consisting of periods of $S^{\Delta}$, has already been shown to be significant for the demonstration of several other phenomena of interest in the study of stimulus control. Jenkins and Harrison (1960) found that a flat generalization gradient was generated on a frequency dimension unless subjects were given discrimination training. Perkins (1953) showed that discrimination training was involved in the production of stimulus intensity dynamism effects. Terrace (1964) demonstrated that discrimination training, with errors, was necessary for generalization peak shift. Weiss (1971) endeavored to isolate the function of $S^{\Delta}$ in the discrimination training of a free-operant stimulus-compounding study.

Weiss (1971, Experiment 1) trained four rats on a two-ply multiple VI 30-second VI 120-second schedule where *no* $S^{\Delta}$ contact was permitted. Subjects were even placed into, and removed from, the training chamber with one of the VI $S^D$s, $T + \bar{L}$ or $L + \bar{T}$, present. According to a "composite-stimulus attentional analysis" the "continuum extremes" ($T + L$ and $\bar{T} + \bar{L}$) would each contain an equivalent representation of stimulus elements from each training condition that maintained responding. Therefore, if one assumes that the animal is sampling composite elements during compounding, as is implicit in Weiss' (1969a) analysis, stimulus compounding should yield no difference in the number of responses to the two "continuum extremes." They should both control response averaging. This is precisely what Weiss (1971, Experiment 1) reported. He stated:

> In light of [these] results . . . , taken in the context of other comparable compounding studies cited, it seems safe to conclude that the discriminations afforded by $S^{\Delta}$ training have been integral to previously reported instances of additive summation [p. 396].

These results of Weiss (1971, Experiment 1) are corroborated by findings of Lawson, Mattis, and Pear (1968, Experiment 2). Lawson et al. trained their four rats on a multiple VI 45-second VI 45-second schedule. Light and tone were $S^D$s. During light, responding was maintained by food, and during tone by water. *There was no $S^{\Delta}$.* In testing, tone plus light controlled a response rate that was intermediate between that forthcoming to tone or light presented separately. Lawson et al. interpreted their failure to get additive summation to compounded $S^D$s to the fact that *different* incentives maintained responding in the presence of each. However, in light of what is now known about the importance of discrimination training between extreme and intermediate composite stimuli for additive summation, it seems that the absence of $S^{\Delta}$ training is the more likely reason for the averaging Lawson et al. found to tone plus light in testing (see also Van Houten & Rudolph, 1971).

Qualitatively, the results of Lawson et al. (1968, Experiment 2) and Weiss (1971, Experiment 1) agree quite favorably with the stimulus-sampling theory analysis of stimulus compounding presented by Atkinson and Estes (1963). Atkinson and Estes cautioned that

> it may be unrealistic to assume that background stimulation from the apparatus and surroundings is negligible. In fact, the experimenter may have to count on an appreciable amount of background stimulation, predominantly conditioned to behaviors incompatible with the CR, to prevent "spontaneous" occurrences of the to-be-conditioned response during intervals between presentations of the experimentally controlled stimuli [p. 196].

When the "background stimulation," or the "composite training extreme" in the terminology of the current presentation, is *not* taken into account in the Atkinson-Estes analysis, the probability of a response to the compound presentation of two independently conditioned stimuli would be somewhere intermediate between the probabilities of each presented separately. (In response rate terms, if rate can be translated to probability of occurrence per unit time, this is precisely what occurred in the two experiments just described where subjects did not receive differential training to $\bar{T} + \bar{L}$.) On the other hand, when differential conditioning to the "background stimulation" is taken into account in the Atkinson and Estes model, and the strengths of the responses controlled by the separately conditioned stimuli are not overly disparate, additive summation is predicted to compound presentations. Later in the current paper it will become apparent that this is the only analysis of stimulus compounding to date that makes the necessity of discrimination training *explicit* for the occurrence of additive summation in conditioning.

---

[8] Donald Meltzer, in a personal communication to the author, presented data confirming these predictions.

### *Consideration of Response Cessation and Nonreinforcement Consequences of $S^\Delta$*

$S^\Delta$ has been an intrinsic part of the training procedures of those free-operant experiments demonstrating additive summation to compounded $S^D$s. Weiss (1971) has shown that it is probably necessary if summation is to be observed, and Weiss (1969a) analyzed its stimulus control function along the composite dimension he postulated to underlie additive summation. But it must be recognized that $S^\Delta$ accomplishes several things behaviorally. In addition to its extinction signal function, $S^\Delta$ comes to control nonresponding. Weiss (1971, Experiment 2) sought to isolate nonreinforcement from response reduction effects of $S^\Delta$ in additive summation designs.

Weiss trained two groups of animals on a three-ply multiple schedule generating similar rates to each respective component between groups. Tone and light were *each* $S^D$s for one of two VI schedules for both groups. The absence of these stimuli, light-out no-tone ($\bar{T} + \bar{L}$) signaled extinction for one group ($S^\Delta$ training) and a DRO (differential-reinforcement-of-behavior-other-than-bar-pressing) food reinforcement schedule for the second group (DRO training). In this experimental design $S^\Delta$ and DRO both controlled response cessation to $\bar{T} + \bar{L}$, but, while $S^\Delta$ was also correlated with nonreinforcement, DRO reinforcement density was maintained intermediate between that controlled by the two VI $S^D$s. Therefore, response cessation was isolated from nonreinforcement during $\bar{T} + \bar{L}$, and the role of differential response rates to the composite-stimulus extreme and VI $S^D$s in summation designs was investigated independently of differences in reinforcement density to the extent that response cessation controlled by DRO and extinction are dynamically similar.

Total test responses were not changed by this manipulation, but rather the proportion of these responses emitted to the compound and the "single" $S^D$s by $S^\Delta$ and DRO subjects. Although both DRO and $S^\Delta$ subjects summated to the $T + L$ compound, DRO subjects emitted a significantly smaller proportion of their total test responses to this compound than did subjects of the $S^\Delta$ group. Therefore, both response cessation and nonreinforcement consequences of $S^\Delta$ appear to be factors determining distribution of total responses during stimulus compounding. If merely the DRO group had been run this multiple causation might have been overlooked, and response cessation or reduction to one of the composite extremes might have been considered the only variable contributing to additive summation.

## SUPPRESSIVE SUMMATION

Suppressive summation is observed when a lesser response is controlled by a compound stimulus than by either of the independently conditioned components presented individually. It was reported by Miller (1969), Reberg and Black (1969, Experiment 1), and Van Houten, O'Leary, and Weiss (1970) in a conditioned suppression paradigm. Miller (1969) and Van Houten et al. (1970) stabilized rats on a VI base line correlated with the absence of tone and light ($\bar{T} + \bar{L}$). Subsequently they made tone ($T + \bar{L}$) and light ($L + \bar{T}$) a signal for unavoidable shock, producing conditioned suppression to each separately. In testing they observed summation of conditioned suppression; that is, a lower rate was forthcoming to the tone and light ($T + L$) presented simultaneously than to the individual presentation of tone or light. Reberg and Black (1969, Experiment 1) demonstrated the same phenomena with shock signaled by darkness and by noise.

### *Composite-Stimulus Control Analysis*

Weiss and Emurian (1970) demonstrated that suppressive summation could be produced to $\bar{T} + \bar{L}$ as well as $T + L$. They reversed the composite extreme associated with the shock-free VI reinforcement contingency to $T + L$, while maintaining $T + \bar{L}$ and $L + \bar{T}$ as CSs, and got less responding to $\bar{T} + \bar{L}$ in testing than to either of the CSs independently presented. Weiss and Emurian (1970) therefore successfully applied the composite-stimulus control model that Weiss (1969a) formulated for additive summation designs to predictions in suppressive summation. They stated:

> Summation of conditioned suppression will be observed to the stimulus complex representing either extreme of a two-modality composite continuum, when "single" stimulus presentations are shock correlated, if the opposite extreme is conditioned as the VI baseline. Therefore the [composite] continuum [in suppressive summation paradigms] is symmetrical with the functions of the two extremes interchangeable [Weiss & Emurian, 1970, p. 208].

Kamin (1963) and Rescorla and Solomon (1967) have suggested that the conditioned suppression phenomenon offers a sensitive measure of Pavlovian conditioning. Therefore, by demonstrating that the "composite-stimulus control" model applies to suppressive summation controlled by conditioned suppression, Weiss and Emurian (1970) appear to have indirectly supported Weiss' (1969a) suggestion that the composite-stimulus analysis could be applied to classical as well as free-operant summation designs. The model has never been tested in a conventional classical conditioning experiment.

## ANALYSIS OF RATE RELATIONS IN ADDITIVE AND SUPPRESSIVE SUMMATION PARADIGMS

Weiss and Emurian (1970) pointed out that in the midst of the many procedural differences between those free-operant experiments yielding additive and those yielding suppressive summation there is an underlying similarity. Response rates are systematically ranked, in opposite directions, over the symmetrical composite dimension in both instances. For example, in an additive summation study the training composite extreme ($\bar{T} + \bar{L}$) would control response cessation, the intermediate composite stimuli ($T + \bar{L}$ or $L + \bar{T}$) would maintain responding, and summation would be forthcoming to the composite opposite to that employed in training ($T + L$). In suppressive summation, the training extreme ($\bar{T} + \bar{L}$) maintains responding, there is a reduction in rate to the intermediate composite stimuli, and $T + L$ produces the lowest rate in testing.

> Therefore what might be the necessary condition for summation [on a multimodal composite dimension], whether additive or suppressive, is differential rates controlled by extreme and intermediate composite stimuli in training. The relation of these rates to each other, other variables held constant, could determine the direction of summation to the composite extreme opposite to that employed in training. [Weiss & Emurian, 1970, p. 209].

However, an incentive difference does exist between those studies reporting additive and those reporting suppressive summation. In those studies reporting additive summation responding to intermediate composites has been maintained by food as well as shock. When suppressive summation has been reported, suppression during training to the intermediate composites has *only* been controlled by a CS signaling an unavoidable shock in a conditioned suppression paradigm, punishment for responding (Van Houten & Rudolph, 1971), or shock avoidance (Weiss & Wiltz, 1972). The rate-relation approach to the analysis of summation would be strengthened if suppressive summation was forthcoming from a subject whose differentiated behaviors to the respective values along the composite continuum were solely controlled by positive reinforcement. A subject has recently yielded these data in the author's laboratory.

A terminal training cumulative record of a rat trained on a multiple (VI 60-second) (conjunctive VI-60-second DRL 13-second) schedule is presented in Figure 1. During $\bar{T} + \bar{L}$, a VI 60-second food reinforcement schedule was in effect. When a tone ($T + \bar{L}$) or a light ($L + \bar{T}$) was present, a DRL 13-second sched-

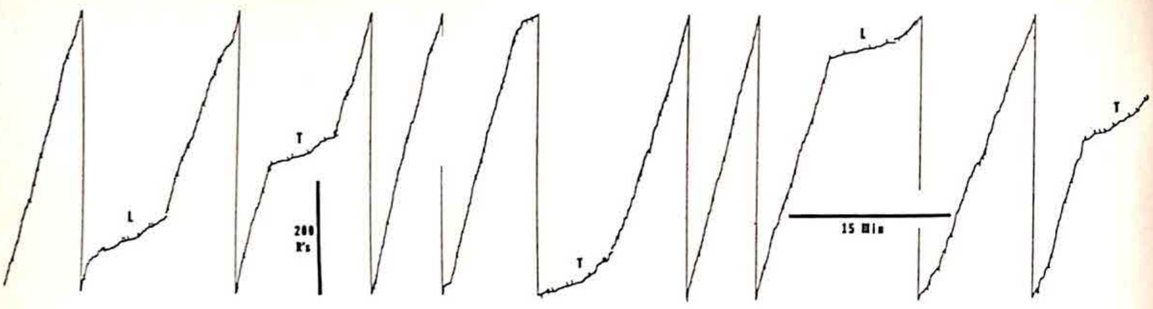


FIG. 1. Terminal cumulative record of rat trained on multiple VI 60-second (VI 60-second DRL 13-second) schedule where the VI 60-second component was signaled by light-out no-tone and the conjunctive VI 60-second DRL 13-second component was correlated with tone (T) or light (L). The latter periods are identified on the record by depression of the cumulative response pen. Slash marks by this pen, whether upward or downward, identify reinforcements.

ule was superimposed on the underlying VI 60-second contingency. This meant that when a reinforcement was made available on the VI contingency, a response would produce it *only* if it occurred at least 13 seconds after the preceding response. In this animal the contingency described maintained a high response rate during $\bar{T} + \bar{L}$ with responding appreciably reduced during $T + \bar{L}$ and $L + \bar{T}$. In more general terms, responding to the training composite stimulus extreme was higher than that controlled by the intermediate composites.

The results of the 12 block-randomized test replications, similar in format to that performed by Weiss and Emurian (1970) when they reported the summation of conditioned suppression, are presented in Table 2. Tone plus light $(T + L)$, $T + \bar{L}$, and $L + \bar{T}$ were each presented 1 minute per block, separated by 3-minute $\bar{T} + \bar{L}$ periods. Only the VI 60-second contingency was effective during the test. Table 2 indicates that suppressive summation was clearly evidenced to $T + L$ over the test. $T + \bar{L}$ and $L + \bar{T}$ controlled approximately one-third the response output of $\bar{T} + \bar{L}$, while $T + L$ controlled roughly one-half as many responses as $T + \bar{L}$ or $L + \bar{T}$. Although this evidence must be viewed as merely supportive of the general theme of the rate-relation analysis of suppressive and additive summation, rather than confirmatory, these preliminary data do lend substantive, and incentive, generality to that interpretation. However, Weiss (1971) cautioned that reinforcement density relations between the extreme and intermediate points along the composite continuum, the manner in which the differential rates between these stimuli are conditioned, as well as the terminal rate relations, could contribute to the control exerted by the compound stimulus represented by the composite extreme opposite to that employed in training. Moreover, recent evidence indicates that even considering all these factors might be insufficient when predicting the distribution of responses between single and compound stimuli in testing. Wiltz[9] just completed a dissertation under the author's supervision in which additive summation was clearly evidenced to combined stimuli $(T + L)$ during extinction *and* maintenance tests, when responding produced food as well as mild

TABLE 2

TOTAL TEST RESPONSES

| Test replications | Stimulus condition | | | |
|---|---|---|---|---|
| | $\bar{T} + \bar{L}$[a] | $T + \bar{L}$ | $L + \bar{T}$ | $T + L$ |
| 1–3 | 141 | 56 | 39 | 42 |
| 5–6 | 208 | 86 | 50 | 28 |
| 7–9 | 179 | 56 | 67 | 25 |
| 10–12 | 186 | 57 | 56 | 26 |
| Total | 714 | 255 | 212 | 121 |

[a] Represents the mean of the 1-minute periods preceding $T + \bar{L}$, $L + \bar{T}$, and $T + L$ stimulus presentations of each test replication totaled of successive blocks of three replications.

---

[9] R. A. Wiltz, Jr. Combining discriminative stimuli signaling punishment. Unpublished doctoral dissertation, The American University, Washington, D. C., 1972.

punishment on independent VI schedules during $T + \bar{L}$ and $L + \bar{T}$, if $\bar{T} + \bar{L}$ controlled response cessation from early training. However, if $\bar{T} + \bar{L}$ first maintained a higher rate than $T + \bar{L}$ and $L + \bar{T}$, when it was subsequently conditioned to control a lower rate than either of these stimuli, additive summation was not forthcoming to $T + L$. It therefore seems, at least in certain schedule situations where summation might be "costly," that the *history* of the rate relations between extreme and intermediate composite stimuli, as well as the immediate rate relations just prior to testing, must be taken into account in a comprehensive analysis of stimulus control. The stimulus-compounding assay appears uniquely sensitive and useful in this analysis.

The rate-relation interpretation of summation in stimulus compounding could mean that compounding stimuli associated with the same schedule, VI 60-second for instance, could under one circumstance produce additive and under another suppressive summation. In fact Weiss and Wiltz (1972) have already demonstrated that compounding $S^D$s controlling free-operant avoidance on an RS 25-second schedule could produce suppressive as well as additive summation as a function of the rate and schedule relations between extreme and intermediate composite stimuli in training. The possible implications of such findings to a comprehensive understanding of "excitatory" and "inhibitory" mechanisms in conditioning as identified by compounding assays are discussed in a later section of this paper.

## AVERAGING

Averaging might be said to occur during stimulus compounding if the response controlled by the compound is somewhere intermediate between the response outputs controlled by the independently conditioned stimuli composing the compound. Pavlov often used the degree of response reduction to CS+ on a summation test to measure the inhibitory strength of a stimulus (Rescorla, 1969a). This measurement technique was necessary because on a simple behavioral level the conditioned tendency not to make a response often cannot be distinguished from the absence of excitation.

TABLE 3

LEPORSKY'S DATA SHOWING ALGEBRAIC SUMMATION OF CONDITIONED REFLEXES

| Time | Stimulus during one minute | Drops of saliva |
|---|---|---|
| 1:40 p.m. | Tone | 21 |
| 2:40 p.m. | Rotating object + tone + flash | 32 |
| 2:10 p.m. | Rotating object | 23 |
| 2:27 p.m. | Simultaneous application of all three positive conditioned stimuli + tactile stimulus | 9 |
| 2:51 p.m. | Rotating object + tactile object | 0 |

### *Some Classical and Operant Conditioning Experiments*

Pavlov (1927) described an experiment by Leporsky that illustrates the effect of a conditioned inhibitor on positive conditioned reflexes of various strengths, as well as summation of conditioned reflexes paired with separate stimuli. Three alimentary reflexes were independently conditioned to a rotating object, to the flash of several lights, and to a musical tone. A tactile stimulus was firmly established as an inhibitor of the response to each of these stimuli. Table 3 not only shows clear summation of conditioned reflexes but also that the conditioned inhibitor, which could completely inhibit the response when presented simultaneously with any one of the positive stimuli, could only partially inhibit the reflex reaction evoked by all three positive conditioned stimuli acting simultaneously. This study clearly demonstrates the algebraic summation of response strengths.

Szwejkowska and Konorski (1959) conditioned two positive CSs—bell ($S_1$) and bubbling water ($S_2$)—while never pairing two other CSs with food—metronome ($S_3$) and whistle ($S_4$). The coincidence of positive and negative CSs ($S_1 + S_3$, $S_2 + S_3$, $S_1 + S_4$, and $S_2 + S_4$) usually produced a salivary effect in the range between the full effect to a positive CS and the minimal response to the negative CS. (Strangely, the simultaneous presentation of both positive CSs produced a CR indistinguishable from either presented separately.)

Cappell, Herring, and Webster (1971), Hammond (1967), Reberg and Black (1969, Experiment 2), and Rescorla (1969b) demonstrated that a CS− (non-shock-associated stimulus) can reduce suppression controlled by a CS+

(shock-associated stimulus) in a conditioned suppression paradigm when CS+ and CS− are presented simultaneously. The CS− is interpreted by these authors as inhibiting the conditioned emotional responses (CERs) elicited by the CS+, thereby reducing the response suppression produced by these CERs.[10]

The results of several operant experiments employing compounding tests similar to Pavlov's are interpreted to indicate that a negative stimulus ($S^\Delta$) exerts inhibitory control. Cornell and Strub (1965) employed differentially located light stimuli as $S^D$s (signaling food availability on a VI schedule) or $S^\Delta$ (signaling extinction) in free-operant training. During compounding tests, compound presentations of $S^D$ and $S^\Delta$ lights led to response rates that were between those controlled by the separate presentations of $S^D$ and $S^\Delta$ lights, indicating, according to Cornell and Strub, the "inhibitory" effect of $S^\Delta$.

Brown and Jenkins (1967) recently performed an experiment that specifically controlled for Skinner's (1938) criticisms of the classical conditioning experiments on which Pavlov (1927) originally based the concept of inhibition. They demonstrated:

> A stimulus (tone) that has become a signal for no responding when paired with one excitatory stimulus (key color used in conjunction with a go/no-go auditory discrimination) also serves as a signal for not responding when it is combined with another excitatory stimulus (key color used for the transfer test) that is clearly discriminated from the one employed in the original training. . . . [Since] extinction of responding to the combination of tone with one key color did not cause loss of response to the second key color, . . . generalized or secondary extinction cannot account for the result [Brown & Jenkins, 1967, p. 262].

The foregoing results fairly well establish through stimulus compounding techniques the functioning of what are customarily thought of as "inhibitory" mechanisms. Nevertheless, Weiss (1967) pointed out:

> When stimulus compounding produced additive summation, the separately conditioned stimuli had each been previously associated with reinforcement and maintained roughly similar, if not identical, response rates or magnitudes. . . . To yield response averaging though, one of the stimuli compounded was reinforcement-associated and maintained responding, while the other signalled extinction . . . and produced a negligible response. . . . Therefore, either the absence of reinforcement to one of the stimuli, or the concomitant low response rate, could be responsible for response averaging observed in these studies. *Response and reinforcement variables are confounded in the comparison of additive summation and response averaging paradigms* [italics added; Weiss, 1967, p. 535].

Weiss (1967) then went on to train animals on a multiple VI 30-second DRL 20-second extinction schedule. After the subjects stabilized, the widely different rates controlled by the VI and DRL $S^D$s satisfied, or surpassed, the rate difference requirements that appeared necessary for free-operant averaging to compounded $S^D$s. On the other hand, the reinforcement densities received during the VI and DRL $S^D$s in training were similar to those associated with independently conditioned $S^D$s that produced additive summation when compounded. Therefore, reinforcement and rate factors were separated.

When the high-rate VI and low-rate DRL training $S^D$s were compounded, they controlled an overall response output that was somewhere intermediate between that produced by these $S^D$s individually presented. In general the same relationship held when generalized stimuli were compounded. Therefore Weiss (1967) concluded that "one of the stimuli compounded does not have to be associated with extinction for response averaging to occur [p. 540]." Widely discrepant rates appear to be sufficient.

---

[10] Results recently reported by Coulter and Weiss (1971) suggested that a stimulus that *previously* served as a CS+ in a conditioned suppression paradigm, but which *currently* signals safety and controls response acceleration, might not act as a "fear inhibitor." Their compound controlled suppressive summation, indicating that the previous conditioning history of, as well as the immediate rate control by, the single stimuli compounded must be appreciated when predicting behavior to a compound stimulus. Significantly, the compounding test behavioral assay appeared to permit the measurement of the residual "fear" still associated with the currently safe CS− of Coulter and Weiss—in spite of the fact that the response acceleration controlled in its presence made it indistinguishable from a CS− that has *always* signaled safety. These results, as well as those presented earlier in the section "Rate Relations," support Konorski's (1967) conclusion "that the kind of primary training of a given stimulus, whether excitatory or inhibitory, determines its later properties: it is difficult to transform an excitatory CS into an inhibitory one . . . , as it is to transform a primarily inhibitory CS into an excitatory one . . . [p. 323]."

### *Interresponse Time Analysis*

Although Weiss (1967) found response averaging when he compounded VI and DRL $S^D$s he did not use these results to attribute "inhibitory" characteristics to the DRL schedule. He rather speculated that the averaging effect was probably due to a competition between the incompatible response repertoires controlled by the two $S^D$s, just as $S^\Delta$ or CS− controls a response repertoire incompatible with that controlled by an $S^D$ or a CS+, respectively. He suggested that the behavioral form of *this* conflict might be more fully appreciated through an analysis of interresponse time (IRT) distributions.

> Bimodal IRTs during compound stimulus presentations could indicate alternating attention between [high-rate] VI- and [low-rate] DRL-associated stimuli, or some intermediate distribution could indicate simultaneous attention to both stimuli [Weiss, 1967, p. 540].

Weiss (1969b, 1972) gathered IRT data during a stimulus-compounding test in which high- and low-rate stimuli were presented simultaneously and individually. A descriptive analysis of response patterning during the three test conditions strongly suggested that averaging total responses to each test condition presents a distorted description of the behavior emitted during the compound. An IRT analysis, as well as examination of test cumulative records, indicated that intermediate response totals to the compound of high- and low-rate $S^D$s were not correlated with an intermediate distribution of responses. The compound distribution appeared to consist of IRTs characterizing the single $S^D$s. The extent to which this finding applies to response averaging produced by the summation of "excitatory" and "inhibitory" response tendencies, when stimuli controlling less disparate response repertoires are compounded, is a ripe area for investigation. Blough (1963), Sewell and Kendall (1965), and Crites, Harris, Rosenquist, and Thomas (1967) have reported that IRT distributions to stimuli removed from the training stimulus on the continuum tested differed from training stimulus-response distributions during generalization *only* with respect to an increased frequency of long IRTs. Migler (1964) has also supplied convincing evidence in a two-manipulandum paradigm that generalization gradients are a statistical artifact of combining mixtures of behaviors over time.

It is obvious that the *likelihood* of finding averaging during compounding depends upon the differences between the response rates to each stimulus. However, more than *these* rate differences seem to be involved. It appears that the discrimination training received between extreme ($\bar{T} + \bar{L}$) and intermediate composite stimuli ($T + \bar{L}$ and $L + \bar{T}$) must be considered in analyzing response averaging as well as additive and suppressive summation. When the composite extreme controls a lower response rate than either of the intermediate composites, it seems reasonable to expect that extremely disparate rates between intermediate composites would be necessary for averaging. In this instance the rate relations between extreme and intermediate composite stimuli would lead to the prediction of additive summation to the composite extreme ($T + L$) opposite to that employed in training, while the incompatibility of rates between intermediate composites would make averaging likely to $T + L$. Averaging and additive summation processes were thus probably operating in opposition to each other when Weiss (1967, 1969b, 1972) reported averaging to compounded tone and light stimuli that customarily controlled rate differences of 25 or greater to one, while $\bar{T} + \bar{L}$ controlled rates lower than either tone or light. (The same thing could probably be said for the Adams and Allen [1971] study where the average rate difference between tone and light in training was 130 responses per minute.) "Incompatible" rates between intermediate composites in training appeared necessary for averaging during compounding in these experiments.

According to the composite stimulus control model discussed earlier, if, in training, different rates are forthcoming to the intermediate composites and no contact is permitted with the composite extreme, or if the rate controlled by the composite extreme is between those controlled by the intermediate composites, response averaging should be forthcoming during compounding. When Weiss (1971, Experiment 1) trained rats on a two-ply multiple VI schedule where tone signaled one VI and light another, and where subjects were allowed *no contact* with $\bar{T} + \bar{L}$, he found clear and consistent response averaging to T + L as well as $\overline{\mathrm{T}} + \overline{\mathrm{L}}$ in testing in spite of the fact that the response rates to tone and light were not characteristically different from those controlled by these stimuli in studies reporting additive summation to T + L. (In these latter studies though [Weiss, 1964; 1969a; 1971 Experiment 2], subjects were given extensive training contact to $\overline{\mathrm{T}} + \overline{\mathrm{L}}$, in whose presence they were taught to *cease* responding.) On the other hand, response averaging reported in conditioned suppression studies is an instance of averaging when the composite stimulus extreme (VI base line) controls a rate between the suppression maintained by the CS+ and the facilitation customarily controlled by CS−. Therefore, due to the different discriminations demanded between the several composite stimulus values in training, whether the free-operant response averaging reported by Weiss (1967, 1969b, 1972) and Weiss (1971, Experiment 1) and the inhibition of suppression studies (Hammond, 1967; Reberg & Black, 1969; Rescorla, 1969b) are dynamically similar is open to question. To be more completely understood, response averaging, just as additive and suppressive summation, will more often have to be considered in the context of the total training schedule controlling the rates to the compounded stimuli.

## "Excitation," "Inhibition," and Stimulus Compounding

Stimulus compounding has been employed as an assay technique to measure the nature of control exerted by independently conditioned stimuli since the early twentieth century. Within the conceptual framework developed by Pavlov, the behavioral resultant to the simultaneous presentation of separately conditioned stimuli could be indicative of the mode of control—"excitatory" or "inhibitory"—conditioned to each. Although these terms are somewhat ambiguous they have, for example, been incorporated as explanatory mechanisms in the comprehensively formalized learning theories of Hull (1943, 1952) and Spence (1956) and are rather loosely used by psychologists in a wide variety of situations.

When compounding two conditioned stimuli results in a greater response than either stimulus presented singly, one observes what is empirically called additive summation. The learning mechanisms postulated to underlie these results would be considered "excitatory" in nature. Hull (1929), in an attempt at a functional interpretation of the conditioned reflex, spoke of it as an additive-subtractive type of reaction in which the number of elements from the original CS complex is the determiner of the intensity of the response. Hull (1929) predicted:

> if two distinct stimuli which have been independently conditioned to a given response are presented together, the intensity of the resulting response is likely to approach the arithmetical sum of the responses to the two stimuli presented separately [p. 502].

In 1943 Hull no longer emphasized the arithmetic sum of the separate responses, but rather the resulting habit strength if the number of reinforcements on all of the separate stimuli has been given consecutively in the presence of one stimulus. Skinner's (1938) law of spatial summation, which reads, "When two reflexes have the same form of response, the response to both stimuli in combination has a greater magnitude and a shorter latency [p. 31]," less formally says the same thing.

Pavlov (1927, p. 76) also suggested that stimulus compounding might be the most direct method of measuring inhibition. When a known exciter of a response is compounded with another stimulus, and the response elicited by the compound is less than that controlled by the exciter alone, the second stimulus is an inhibitor. In addition to predicting the summation of compatible reaction potentials, Hull (1943) also noted that there can be a competition of incompatible reaction potentials that would lead to a less intense response, or no response at all if the antagonistic reaction potentials are momentarily equal. Functionally, Skinner (1957) accepted the same behavioral resultants when he stated:

> it is a well established principle in nonverbal behavior that separate sources of strength are additive. (Since some variables reduce the strength of verbal behavior, the addition must be algebraic) [p. 228].

Therefore, the results of stimulus compounding tests have identified the interaction of "exciters" and of "exciters" and "inhibitors." Following the same line of reasoning, if the simultaneous presentation of two independently conditioned stimuli produces a lesser response than either presented separately, the addition of "inhibition" should be indicated. Most likely this would be observed in classical conditioning when a lesser response was forthcoming when an "exciter" was compounded with two rather than with one "inhibitor." Suppressive summation in operant conditioning has been extensively discussed in a preceding section of this paper.

Within the framework sketched "excitation" and "inhibition" are mutually exclusive processes that work in opposition to one another. However, on the basis of previous research it was predicted that the response rate and reinforcement density relations between the two stimuli compounded, and the composite extreme of a three-ply multiple training schedule, would be the factors determining direction of summation. Within this framework, where compounding precisely the *same* schedule-associated stimuli can lead to additive or suppressive effects, the concepts of "excitation" and "inhibition" take on a definite relational character. This indicates that the stimulus control exerted by a component of a multiple schedule cannot be fully appreciated without due regard to the total schedule context of which it is a part. This is what Pavlov, Hull, and Skinner failed to make explicit when they analyzed stimulus compounding. In terms of Rescorla's (1969a) distinction, compounded stimuli could control tendencies "opposite" to one another, addition or suppression, in testing although they specifically were *each* associated with the *same* particular schedules of reinforcement controlling similar response repertoires. *These results*, not altogether speculative as data presented in prior sections of this paper indicate, *would require a reformulation, at least as far as a stimulus-compounding index of these processes is concerned, of "excitatory" and "inhibitory" mechanisms in conditioning*. However, these results of compounding would agree substantially with Hearst, Besley, and Farthing's (1970) recent redefinition of an inhibitory stimulus "as a stimulus that develops during conditioning the capacity to decrease response strength below the level occurring in its absence [p. 405]." According to this definition, if response strength may be translated to a rate measure, it could be the *rate relations* between components of a multiple schedule that determine inhibitory, and by implication excitatory, stimulus control, just as they can determine those circumstances that produce additive and suppressive summation during compounding. It appears, however, that once these mechanisms are defined in such a fashion that they lose their absolute properties, their conceptual usefulness, from an integrative point of view, is severely diminished. As Konorski (1967) suggested, it might be more useful and parsimonious to think in terms of the interaction of different excitatory mechanisms.

## GENERALIZATION PEAK SHIFT AND SUMMATION: SOME FUNCTIONAL SIMILARITIES

Positive peak shift is observed during a generalization test when maximum responding is controlled by a stimulus removed from the $S^D$ controlling the highest response rate in training in a direction away from that stimulus controlling a lower training rate. If the on–off states of the stimuli employed in those stimulus compounding experiments that report additive summation might be conceptualized as forming a composite continuum extending from the all-off extreme ($\bar{T} + \bar{L}$) through the one-stimulus-on condition ($T + \bar{L}$ or $L + \bar{T}$), to the all-on extreme ($T + L$), a functional comparison of summation to peak shift would be facilitated (see the stimulus identification row of Table 1).

Wolf (1963) and Weiss (1964) were the first to report that subjects summated additively to $T + L$ in testing when responding in training was extinguished to $\bar{T} + \bar{L}$ and maintained through reinforcement to $T + \bar{L}$ and $L + \bar{T}$. On the composite continuum, $T + L$ can be seen as further removed from $\bar{T} + \bar{L}$ than either $T + \bar{L}$ or $L + \bar{T}$. This led Wolf to compare the response enhancement effect of combined $S^D$s to generalization peak shift. He postulated that the composite stimulus dimension employed in summation experiments might be functionally equivalent to the unitary dimensions of the peak shift studies.

Subsequent to Wolf's suggestion, evidence has been reported that is supportive of the peak shift-summation comparison. Response enhancement effects along the ratio scale continua employed in peak shift studies, as well as along the composite continuum postulated to underlie summation, appear to be controlled by similar variables. These include (*a*) intradimensional discrimination training that produces rate differences between stimulus values along the respective continua, with the *possibility* that contrast must be a by-product of that training if either enhancement phenomena is to occur; (*b*) reversibility of function between high- and low-rate training stimuli in positive and negative peak shift as well as additive and suppressive summation; and (*c*) the contribution of both rate and reinforcement differences between training stimuli in determining magnitude of response enhancement for peak shift and summation in testing.

Positive peak shift will occur after discrimination training in which one value on a continuum is reinforced and another extinguished (cf. Hanson, 1959). Either value along the stimulus continuum can serve equally well in either $S^D$ or $S^\Delta$ capacity, with the peak shifted one way or the other as a function of the position of $S^\Delta$ relative to $S^D$. Weiss (1969a) demonstrated this reversibility of function in an additive summation paradigm. Prior to this experiment, additive summation had been reported to $T + L$ when $T + \bar{L}$ and $L + \bar{T}$ were each $S^D$s for a VI food reinforcement contingency, while $\bar{T} + \bar{L}$ signaled extinction ($S^\Delta$). $T + \bar{L}$ and $L + \bar{T}$ were *also* the VI $S^D$s when summation was reported to $\bar{T} + \bar{L}$ by Weiss (1969a), but in this instance $S^\Delta$ was signaled by $T + L$. These results confirmed that the composite continuum was behaviorally symmetrical in additive summation, with the functions of the two extremes interchangeable in much the same way that 550 or 570 millimicrons (m$\mu$), for example, can be employed as $S^\Delta$ to produce peak shift. Further, it is well documented that intradimensional discriminative training is necessary for peak shift just as Weiss (1971, Experiment 1) suggested that it was essential to additive summation. His subjects, who did not receive discrimination training between *different* values along the composite continuum, defined in terms of the number of composite elements "on," failed to summate additively to compound stimulus presentations.

If additive summation might be conceptualized as having attributes in common with positive peak shift, suppressive summation might be compared to negative peak shift. Guttman (1965) and Hendry, Switalski, and Yarczower (1969) reported negative peak shift. In Guttman's experiment, responding in training was not completely eliminated to $S^\Delta$ but was merely maintained at a lower level to $S^\Delta$ than $S^D$. In testing, the stimulus that controlled the lowest response output was not $S^\Delta$ but a stimulus value removed from $S^\Delta$ in a direction away from $S^D$.

During suppressive summation the composite extreme opposite to that which maintains responding controls a lower response output than either of the intermediate composites that control suppression in training, just as a stimulus further removed from $S^D$ than $S^\Delta$ controls a lower response output than $S^\Delta$ in negative peak shift. Guttman (1965) produced negative peak shift when 550 m$\mu$ was $S^D$ and 560 m$\mu$ was $S^\Delta$, as well as when the opposite stimulus-schedule combination was employed. Similarly, Weiss and Emurian (1970) demonstrated that suppressive summation could be controlled by either extreme of their composite continuum after the appropriate training conditions.

That rate relations between the components of a multiple schedule might be of fundamental importance in determining stimulus control has recently been demonstrated by Yarczower, Dickson, and Gollub (1966) and Terrace (1966, 1968). Yarczower et al. (1966) systematically varied response rate and reinforcement density relations during differential intradimensional training between components of a multiple schedule. Their manipulations led these authors to conclude that their "data imply that response rates at the end of training, rather than reinforcements *per se*, determine the characteristics of the generalization gradient [Yarczower et al., 1966, p. 631]." When rates between multiple-schedule components were different, the peaks of their generalization gradients were shifted away from the lower rate stimulus. On the other hand, Yarczower et al. showed that as long as the rates in the two components were equal, the generalization peaks were not shifted even when reinforcement densities were not equal. More recently Yarczower, Gollub, and Dickson (1968) have suggested that reinforcement density *and* rate relations between the components of a multiple schedule might interact in determining gradient shape. Weiss (1971, Experiment 2) demonstrated that these same variables appear to be operating to determine magnitude of additive summation.

Terrace (1968) affirmed the findings of Yarczower et al. (1966) while extending the rate relation analysis to behavioral contrast (cf. Reynolds, 1961) in addition to peak shift. Contrast is observed when, after discrimination training on a multiple schedule, the rate in the presence of one component that remains unchanged moves in a direction opposite to that produced by a change in the other schedule. Terrace (1968), from the results of a variety of experiments that produced differential responding to the components of a multiple schedule, concluded:

> that a reduction of the rate of responding to one of the two alternating discriminative stimuli is sufficient to produce contrast. The present results indicate also that a sufficient condition for the occurrence of the peak shift is a reduction in the rate of responding to one of two alternating discriminative stimuli . . . [with the peak shifted] away from the stimulus in the presence of which the rate of responding was reduced [1968, p. 739].

Rate reduction to a composite extreme has not yet explicitly been investigated in stimulus compounding yielding additive summation. Usually, though, when differential conditioning starts from the beginning of training, rate reduction to one of the components is implicit during acquisition of the discrimination. In fact, Weiss (1971, Experiment 2) documented this reduction. Therefore, although not explicitly investigated, rate reduction to the composite extreme employed in training has probably been implicit in those studies reporting additive summation. *Thus, to the extent the evidence presented supports the contention that peak shift and summation possess some underlying functional similarities, the postulated operation of rate differences between training extreme and intermediate composite stimuli in determining stimulus control properties within compounding experiments is supported by the importance of that variable to peak shift.*

## CONCLUSION

Research interest in the area of stimulus compounding has increased substantially in the last several years. Additive summation is now a well-established phenomenon, with the free-operant research in particular indicating that it is not schedule or incentive-specific. Rather, additive summation appears to be dependent upon discrimination training that leads to a *higher rate*, or response output, in the presence of *each* of the stimuli to be compounded than in their *simultaneous* absence. Research to date has customarily maintained a minimal response output to this latter condition by programming extinction in its presence. However, responding has been maintained to the individual stimuli subsequently compounded by VI, FI, DRL and signaled as well as Sidman avoidance contingencies in the various free-operant studies surveyed. The classical conditioning experiments yielding additive summation were more limited in their scheduling, although a variety of reinforcers were used as unconditioned stimuli.

Unfortunately, attempts to predict the response output to a compound stimulus presentation as a function of the responses controlled by the independently conditioned stimuli composing it have not been overly successful in the literature of additive summation. The results have ranged from additive summation effects, just barely meeting the minimum requirements defined earlier in this paper, to several studies that fit a simple additive model, to others that report *substantially* greater responding to the compound than to the outputs controlled by the single stimuli combined. This variability is not surprising in view of the fact that it is just recently that some of the dynamics underlying additive summation have begun to be made explicit (Weiss, 1969, 1971; Weiss & Emurian, 1970), and a variety of training and testing procedures have been employed by different investigators. In particular, test length might be a significant factor in determining the magnitude of additive summation since several investigators employing extinction tests have specifically stated that the compound test condition continued to control responding when responding to each of the single stimuli composing it was completely extinguished.

Suppressive summation has only been reported in the free-operant literature, and then only when response suppression to the individual stimuli comprising the compound was controlled by punishment or unavoidable signaled shock. Data presented in this paper suggested that the phenomenon is not, however, incentive-specific, but rather that it could also be observed when the rate reductions in training to the single stimuli compounded was controlled by food-maintained schedules. This finding lends further weight to the contention that suppressive summation is dependent upon, within limits, discrimination training that leads to a *lower rate*, or response output, in the presence of *each* of the stimuli to be compounded than in their *simultaneous* absence.

Examining the influence of discrimination training on response distributions during stimulus compounding led to the development of an empirically substantiated composite-stimulus control model applicable to additive and suppressive effects in both free-operant and classical conditioning. A symmetrical composite-stimulus continuum, defined by the on–off states of the stimuli controlling behavior in training, was seen to be common to both types of summation, with the functions of the all-on and all-off continuum extremes interchangeable. Moreover, additive and suppressive summation seem to be specifically determined by the relations between response rates and reinforcement densities conditioned to particular stimulus values along the composite training continuum—with the most recent evidence indicating the importance of the *history* of these relations, as well as those conditioned immediately prior to testing, in determining the distribution of responses during a compounding test. Since, according to this formulation, compounding the *same* schedule-associated stimuli might lead to additive summation under one set of circumstances, suppressive summation under another, or even averaging in a third complex schedule context, *absolute* properties cannot be attributed to these schedule-associated behaviors without due regard to the total schedule context of which they are a part. This indicates that the continued exploration of those factors determining additive and suppressive effects in compounding might further help in clarifying, and placing in contextual perspective, what are currently thought of as "excitatory" and "inhibitory" mechanisms in learning—thereby even stimulating the development of more advanced unifying principles of behavioral control. For the most part, stimulus compounding research to date has focused on two-element composite control. As higher-element composites are explored, more complex interactions will inevitably be revealed by this powerful behavioral assay.

## REFERENCES

Adams, D. L., & Allen, J. D. Compound stimulus control by discriminative stimuli associated with high and moderate response rates. *Journal of the Experimental Analysis of Behavior*, 1971, **16**, 201–205.

Atkinson, R. C., Calfee, R. C., Sommer, G. R., Jeffrey, W. E., & Shoemaker, R. A test of three models for stimulus compounding with children. *Journal of Experimental Psychology*, 1964, **67**, 52–58.

Atkinson, R. C., & Estes, W. K. Stimulus sampling theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. II. New York: Wiley, 1963.

Baker, T. W. Properties of compound conditioned stimuli and their components. *Psychological Bulletin*, 1968, **70**, 611–625.

Blough, D. S. Interresponse time as a function of continuous variables: A new method and some data. *Journal of the Experimental Analysis of Behavior*, 1963, **6**, 237–246.

Brown, P. L., & Jenkins, H. M. Conditioned inhibition and excitation in operant discrimination learning. *Journal of Experimental Psychology*, 1967, **75**, 255–266.

Cappell, H. D., Herring, B., & Webster, C. D. Discriminated conditioned suppression: Further effects of stimulus compounding. *Psychonomic Science*, 1970, **19**, 147–149.

Cornell, J. M., & Strub, H. A technique for demonstrating the inhibitory function of $S^{\Delta}$. *Psychonomic Science*, 1965, **3**, 25–26.

Coulter, W. R., & Weiss, S. J. Suppressive summation controlled by compounding a shock-associated and a shock-extinguished CS. *Proceedings of the 79th Annual Convention of the American Psychological Association*, 1971, **6**, 711–712.

Crites, R. J., Harris, J. T., Rosenquist, H., & Thomas, D. P. Response patterning during stimulus generalization in the rat. *Journal of the Experimental Analysis of Behavior*, 1967, **10**, 165–168.

Emurian, H. H., & Weiss, S. J. Compounding discriminative stimuli controlling free-operant avoidance. *Journal of the Experimental Analysis of Behavior*, 1972, **17**, 249–256.

Estes, W., & Skinner, B. Some quantitative properties of anxiety. *Journal of Experimental Psychology*, 1941, **29**, 390–400.

Grings, W. W., & O'Donnell, D. E. Magnitude of response to compounds of discriminated stimuli. *Journal of Experimental Psychology*, 1956, **52**, 354–359.

Grings, W., Uno, T., & Fiebiger, J. Component to compound stimulus transfer. *Psychonomic Science*, 1965, **3**, 63–64.

Guttman, N. Effects of discrimination formation on generalization measured from a positive-rate baseline. In D. Mostofsky (Ed.), *Stimulus generalization*. Stanford, Calif.: Stanford University Press, 1965.

Hammond, L. J. A traditional demonstration of the active properties of Pavlovian inhibition using differential CER. *Psychonomic Science*, 1967, **9**, 65–66.

Hanson, H. M. Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, 1959, **58**, 321–334.

Hearst, E., Besley, S., & Farthing, G. W. Inhibition and the stimulus control of operant behavior. *Journal of the Experimental Analysis of Behavior*, 1970, **14**, 373–409.

Hendry, D., Switalski, R., & Yarczower, M. Generalization of conditioned suppression after differential training. *Journal of the Experimental Analysis of Behavior*, 1969, **12**, 799–806.

Hull, C. L. A functional interpretation of the conditioned reflex. *Psychological Review*, 1929, **36**, 498–511.

Hull, C. L. Exploration in the patterning of stimuli conditioned to the GSR. *Journal of Experimental Psychology*, 1940, **27**, 95–110.

Hull, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

Hull, C. L. *A behavior system*. New Haven: Yale University Press, 1952.

Jenkins, H. M., & Harrison, R. H. Effect of discrimination training on auditory generalization. *Journal of Experimental Psychology*, 1960, **59**, 246–253.

Kamin, L. J. Backward conditioning and the conditioned emotional response. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 517–519.

Kimble, G. H. *Hilgard and Marquis' conditioning and learning*. New York: Appleton-Century Crofts, 1961.

Konorski, J. *Conditioned reflexes and neuron organization*. (Trans. by Stephen Garry) Cambridge, England: Cambridge University Press, 1948. (Reprinted New York: Hafner 1968).

Konorski, J. *Integrative activity of the brain*. Chicago: University of Chicago Press, 1967.

Lawson, R., Mattis, P. R., & Pear, J. J. Summation of response rates to discriminative stimuli associated with qualitatively different reinforcers. *Journal of the Experimental Analysis of Behavior*, 1968, **11**, 561–568.

Meltzer, D., & Freeman, B. Maintenance of response summation under conditions of minimum intensity. *Psychonomic Science*, 1971, **22**, 287–289.

Migler, B. Effects of averaging data during stimulus generalization. *Journal of the Experimental Analysis of Behavior*, 1964, **7**, 303–307.

Miller, L. Compounding of pre-aversive stimuli. *Journal of the Experimental Analysis of Behavior*, 1969, **12**, 293–299. (a)

Miller, L. Stimulus compounding with an instrumental avoidance response. *Psychonomic Science*, 1969, **16**, 46–47. (b)

Miller, L. Compounding of discriminative stimuli from the same and different modalities. *Journal of the Experimental Analysis of Behavior*, 1971, **16**, 337–342.

Miller, L., & Ackley, R. Summation of responding maintained by fixed-interval schedules. *Journal of the Experimental Analysis of Behavior*, 1970, **13**, 199–203.

Miller, L., & Price, R. D. Stimulus compounding and response summation with an instrumental running response. *Psychonomic Science*, 1971, **22**, 52–53.

Pavlov, I. P. *Conditioned reflexes*. (Trans. by G. V. Anrep) Oxford: Oxford University Press, 1927.

Perkins, C. C., Jr. The relation between conditioned stimulus intensity and response strength. *Journal of Experimental Psychology*, 1953, **46**, 225–231.

Razran, G. Studies in configural conditioning: I. Historical and preliminary experimentation. *Journal of General Psychology*, 1939, **21**, 307–330.

Razran, G. Empirical codifications and specific theoretical implications of compound-stimulus conditioning: Perception. In W. F. Prokasy (Ed.), *Classical conditioning*. New York: Appleton-Century-Crofts, 1965.

Reberg, D., & Black, A. H. Compound testing of individually conditioned stimuli as an index of excitatory and inhibitory properties. *Psychonomic Science*, 1969, **17**, 30–31.

Rescorla, R. A. Pavlovian conditioning and its proper control procedures. *Psychological Review*, 1967, **74**, 71–80.

Rescorla, R. A. Pavlovian conditioned fear in Sidman avoidance learning. *Journal of Comparative and Physiological Psychology*, 1968, **65**, 55–60.

Rescorla, R. A. Pavlovian conditioned inhibition. *Psychological Bulletin*, 1969, **72**, 77–94. (a)

Rescorla, R. A. Conditioned inhibition of fear resulting from negative CS-US contingencies. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 504–509. (b)

Rescorla, R., & Solomon, R. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychological Review*, 1967, **74**, 151–182.

Reynolds, G. S. Behavioral contrast. *Journal of the Experimental Analysis of Behavior*, 1961, **4**, 57–71.

Riess, D. Pavlovian phenomena in conditioned acceleration—Stimulus summation. *Conditional Reflex*, 1969, **4**, 257–264.

Schoeffler, M. S. Probability of response to compounds of discriminated stimuli. *Journal of Experimental Psychology*, 1954, **48**, 323–329.

Sewell, W. R., & Kendall, S. B. A note on interresponse time distributions during generalization testing. *Psychonomic Science*, 1965, **3**, 95–96.

Skinner, B. F. *The Behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts, 1938.

Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

Sidman, M., Herrnstein, R. J., & Conrad, D. G. Maintenance of avoidance behavior by unavoidable shocks. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 553–557.

Spence, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.

Szwejkowska, G., & Konorski, J. The influence of the primary inhibitory stimulus upon the salivary effect of an excitatory conditioned stimulus. *Acta Biologiae Experimentalis*, 1959, **19**, 161–174.

Terrace, H. S. Wavelength generalization after discrimination learning with and without errors. *Science*, 1964, **144**, 78–80.

Terrace, H. S. Stimulus control. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application*. New York: Appleton-Century-Crofts, 1966.

Terrace, H. S. Discrimination learning, the peak shift, and behavioral contrast. *Journal of the Experimental Analysis of Behavior*, 1968, **11**, 727–741.

Timofeev, N. N. Comparative physiological studies on exteroceptive and interoceptive conditioned reflexes. (Reprinted: *The central nervous system and human behavior*. Washington, D. C.: National Institutes of Health, 1959).

Van Houten, R., O'Leary, K. D., & Weiss, S. J. Summation of conditioned suppression. *Journal of the Experimental Analysis of Behavior*, 1970, **13**, 76–82.

Van Houten, R., & Rudolph, R. Summation of punishment suppression. *Journal of the Experimental Analysis of Behavior*, 1971, **15**, 117–121.

Weiss, S. J. Summation of response strengths instrumentally conditioned to stimuli in different sensory modalities. *Journal of Experimental Psychology*, 1964, **68**, 151–155.

Weiss, S. J. Free-operant compounding of variable-interval and low-rate discrimination stimuli. *Journal of the Experimental Analysis of Behavior*, 1967, **10**, 535–540.

Weiss, S. J. Attentional processes along a composite stimulus continuum during free-operant summation. *Journal of Experimental Psychology*, 1969, **82**, 22–27. (a)

Weiss, S. J. Response distributions during free-operant compounding of high- and low-rate discriminative stimuli. *Proceedings of the 77th Annual Convention of the American Psychological Association*, 1969, **4**, 827–828. (b) (Summary)

Weiss, S. J. Discrimination training and stimulus compoundings: Consideration of non-reinforcement and response differentiation consequences of $S^{\Delta}$. *Journal of the Experimental Analysis of Behavior*, 1971, **15**, 387–402.

Weiss, S. J. Free-operant compounding of high- and low-rate discriminative stimuli: An interresponse time analysis. *Learning and Motivation*, 1972, in press.

Weiss, S. J., & Emurian, H. Stimulus control during the summation of conditioned suppression. *Journal of Experimental Psychology*, 1970, **85**, 204–209.

Weiss, S. J., & Wiltz, R. A., Jr. Response enhancement and suppression through compounding discriminative stimuli signaling free-operant avoidance. *Proceedings of the 80th Annual Convention of the American Psychological Association*, 1972, **7**, 757–758.

Wolf, M. M. Some effects of combined $S^D$s. *Journal of the Experimental Analysis of Behavior*, 1963, **6**, 343–347.

Yakovleva, V. V. Individual conditioned stimuli applied in combination and then separately. *Trudy fiziologicheskikh laboratoriĭ akademika I.P. Pavlova*, 1927, **2**, 37–45.

Yarczower, M., Dickson, J. F., & Gollub, L. R. Some effects on generalization gradients of tandem schedules. *Journal of the Experimental Analysis of Behavior*, 1966, **9**, 631–639.

Yarczower, M., Gollub, L. R., & Dickson, J. F. Some effects of discriminative training with equated frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 1968, **11**, 415–423.

(Received February 15, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 3, 209–233

# THE 47,XYY MALE:

## A REVIEW

DAVID R. OWEN[1]

*University of Iowa*

Reports of males with a 47,XYY chromosome complement are reviewed in regard to psychological and physical variables. No clear relationship of external genitalia, bone and joint abnormalities, dental irregularities, skin disorders, and some hormone levels with an XYY constitution has been established. Increased stature, dermatoglyphic alterations, possible neurological involvement, alterations in testicular cells, and some slight cardiac anomaly may, with properly collected data, prove to be predictable from the XYY genome. Data relevant to psychological and social functioning include diagnoses, personality and behavioral descriptions, criminal records, age at first conviction, and psychometric scores on tests of personality and intelligence. No consistent personality or behavioral constellation has been successfully predicted from the XYY complement. Recommendations are made for future researchers in this area of behavior genetics.

The August 26, 1961, *Lancet* carried a letter announcing an observation of the last of the abnormal genomes that could result from a faulty separation of the sex chromosomes during oögenesis and/or spermatogenesis (Sandberg, Koepf, Ishihara, & Hauschka, 1961). That letter constituted the first claim of an observation of a 47,XYY human. Subsequently, reanalysis (Dent, Edwards, & Delhanty, 1963) of a 1960 report (Hayward & Bower) on an abnormal karyotype suggested it was an XYY rather than 22-trisomy as earlier reported.

During the following decade, examples of 47,XYY males have been reported from Australia, Bulgaria, Canada, Denmark, England, Finland, France, Germany, Hungary, Italy, Jamaica, Japan, Mexico, the Netherlands, Poland, Scotland, South Africa, Sweden, and the United States. Recently the XYY sex chromosome constitution has been observed in mice (Cattenach & Pollard, 1969; Evans, Ford, & Searle, 1969).

During the first decade with the XYY male, investigations proceeded up many blind alleys, due primarily to a disregard for proper sampling techniques. Early hopes that the 47,XYY chromosome constitution might lead to a relatively specific set of anomalies, as had 21-trisomy, were soon shown unrealistic. During this struggle to understand the XYY male, studies have proceeded through three phases (Court Brown, 1968). The collection of case reports in which the XYY was discovered rather by accident constituted the first phase. The second phase began with intensive efforts to locate concentrations of XYY males in various subpopulations, especially mental hospitals and prisons. The third phase finds investigators addressing themselves to the real issues: What is the population incidence rate, and what anomalies, if any, can be predicted given a 47,XYY karyotype?

From these studies, a mass of data on men with 47,XYY chromosomes has accumulated. Unfortunately, the bulk of these data is of indeterminate value for a genuine understanding of the implications of an XYY pattern. Most cases reported during the first phase of the research were karyotyped because they possessed a phenotypic attribute that was either unusual or previously noted in an XYY case study. Within the initial phase, selection bias seemed to operate primarily toward genital anomalies. The second phase, with the suggestion that abnormal tendencies toward aggressive behavior might be associated with the XYY, suffered a strong selection bias for the karyotyping of tall, imprisoned subjects. The third phase, with limited selection bias, introduced a considerable danger of Rosenthal effects. Although no long-range data are yet available on 47,XYY males discovered by mass screening programs, these data will become most valuable if kept reasonably free from contamination by expectancies. The oldest of these XYY children is about 4 years now (Valentine, 1969).

---

[1] Requests for reprints should be sent to David R. Owen, who is now at the Division of Psychological Studies, Educational Testing Service, Princeton, New Jersey 08540.

Caution is fully warranted in accepting 47,XYY diagnoses.

> *Accurate* identification of individual chromosomes is at best extremely difficult and, more often than not, impossible. . . . For example, the Y chromosome has been variously regarded as the longest of [Ferguson-Smith, Johnston, & Weinberg, 1960; Patau, 1961], the smallest of [Chu & Giles, 1959], and the same size as [Ford & Hamerton, 1956] small acrocentric chromosomes (21 and 22). More recently, reports [Bender & Gooch, 1961; van Wijck, Tijdink, & Stolte, 1962] have indicated that the Y chromosome varies in size not only from person to person but from cell to cell within the same person [Eggen, 1963, p. 9].

Techniques have presumably improved with additional experience. Nonetheless, subjective judgment often entered into the karyotyping procedures. Recent reports (Anonymous—Pursuit of Y, 1970; Barlow & Vosa, 1970; George, 1970; Pearson & Bobrow, 1970; Pearson, Bobrow, & Vosa, 1970; Vosa, 1970; Zech, 1969) indicate that a fast and reliable technique may have been discovered for identifying Y chromosomes in human interphase nuclei. Staining with quinacrine mustard or quinacrine dihydrochloride (Atebrin) causes the Y chromosome to fluoresce brightly. This may allow much more efficient mass screening for Y chromosome aneuploidies, comparable to staining for sex chromatin in screening for supernumary X chromosomes. Borgaonkar and Hollander (1971) report, however, that no fluorescence could be detected in preparations from patients with small Y chromosomes.

An attempt has been made in the following review to exclude XYY males known to be chromosome mosaics. Also excluded are XYY cases accompanied by any other chromosome anomaly (e.g., Down's syndrome). Cases of apparent 47,XYY females are briefly discussed, but they are excluded from summary data.

## PREVALENCE AND INCIDENCE OF 47,XYY

One of the most widely discussed hypotheses about 47,XYY males is that supernumary Y chromosomes predispose the holder to violent acts of masculine aggression. What began as a reasonable hypothesis—but immediately failed to gain statistical support—was accepted uncritically by an apparently alarming number of practitioners.

Early surveys of sex-chromatin (Barr body) rates disclosed prevalence of around 2% in institutions for male criminals and the "hard-to-manage" (Casey, Segall, Street, & Blank, 1966; Forssman & Hambert, 1963). A follow-up revealed that about one-third of these were 48,XXYY, with the others being 47,XXY. An earlier survey of ordinary mental subnormal males had observed a lower rate of 48,XXYY (Maclean, Mitchell, Harnden, Williams, Jacobs, Buckton, Baikie, Court Brown, McBride, Strong, Close, & Jones, 1962), giving rise to a hypothesized relationship between the extra Y chromosome and aggressive behavior.

An immediate search for partial validity was undertaken in institutions containing individuals detained because of criminal or delinquent behavior. Among 342 mentally subnormal patients at Carstairs Hospital, 9 with 47,XYY were found (Jacobs, Brunton, Melville, Brittain, & McClemont, 1965; Price, Strong, Whatmore, & McClemont, 1966). This prevalence rate was indeed high, and it generated a great deal of enthusiasm. However, offering this finding, striking as it is, as an indication that 47,XYY males are particularly likely to arrive in a hospital for aggressive criminals is premature. Some attempt must be made to determine if a Type I error occurred. That is, could this extremely improbable event (9/342) have been just that—an extremely improbable event?

Table 1 includes a number of attempts to replicate the Carstairs data. Seven of these studies relate to the Carstairs East data on mentally subnormal. Three of these studies represent a population of 2,207 patients, with two 47,XYY cases among the 764 patients karyotyped. At Letchworth Village, 1 XYY subject was found among 4,000 male and female residents by karyotyping 55 boys. A fifth study failed to find any 47,XYY among 330 mentally retarded. A sixth study found 1 among the 32 karyotyped. And a seventh study found 5 among the 21 karyotyped.

TABLE 1

INSTITUTIONAL RATES OF 47,XYY

| Author(s) and year | Population size | Selection criteria | Number karyotyped | Number 47,XYY | Institution |
|---|---|---|---|---|---|
| Mentally subnormal | | | | | |
| Jacobs, Brunton, Melville, Brittain, & McClemont (1965) | 203 | None | 197 | 7 | Carstairs—East |
| Close, Goonetilleke, Jacobs, & Price (1968) | 917 | 5′10″+ | 68 | 2 | Darenth Park |
| Jacobs, Price, Court Brown, Brittain, & Whatmore (1968) | 590 | IQ < 40 | 200 | 0 | Hospital for mentally subnormal |
| | | IQ > 40 | 390 | 0 | |
| Clark, Telfer, Baker, & Rosen (1970) | 700 | 71″+ | 6 | 0 | Mental retardation |
| | | 71″− | 100 | 0 | |
| Shapiro (1970a) | 4,000[a] | 5′9″+ | 55 | 1 | Letchworth Village |
| Price (1969) | ? | 183+ cm. | 21 | 5 | Coleshill Hall and Chelmsley |
| Weiner, Sutherland, Bartholomew, & Hudson (1968) | ? | 72″+; criminal record | 30 | 0 | Mentally retarded |
| | ? | ? | 300 | 0 | Mentally retarded |
| Daly (1969a) | ? | Tall | 32 | 1 | Mentally retarded |
| Mental hospitals | | | | | |
| Jacobs et al. (1968) | 139 | None | 118 | 2 | Carstairs—West |
| Casey, Blank, Street, Segall, McDougall, McGrath, & Skinner (1966) | 680[b] | 6′+ | 50 | 4 | Broadmoor |
| | 1,500±[c] | 6′+ | 30 | 0 | Mentally ill |
| Nielsen (1968) | 440 | 180+ cm. | 42 | 0 | Arhus State Hospital |
| Akesson, Forssman, & Wallin (1968) | 96 | Consecutive admissions | 96 | 3 | Mental hospital |
| Clark et al. (1970) | 1,572 | 71″+ | 72 | 0 | State mental hospital |
| | | 71″− | 87 | 0 | |
| Marinello, Berkson, Edwards, & Bannerman (1969) | ? | 183+ cm. | 76 | 1 | Gowanda State Hospital |
| Parker, Melnyk, & Fish (1969) | ? | 6+; History of aggressive behavior | 10 | 1 | Fairview State Hospital |
| Epilepsy | | | | | |
| Forssman & Akesson (1969) | 320 | 183+ cm. | 14 | 0 | Long-term care |
| Court Brown (1968) | 71 | None | 71 | 1 | Epileptic colony |
| Prisons and criminally insane | | | | | |
| Casey et al. (1966) | 203[d] | 6′+ | 24 | 2 | Nottingham Prison |
| | 942[e] | 6′+ | 50 | 12 | Rampton and Moss Side |
| Goodman, Smith, & Migeon (1967) | 3,000± | 73″+; Volunteers | 100 | 2 | Ohio Penitentiary |
| Welch, Borgaonkar, & Herr (1967) | 464 | 72″+; IQ < 75 | 10 | 0 | Patuxent Institution |
| | | 12 most aggressive; | 12 | 0 | |
| | | 75″+ | 22 | 1 | |
| Wiener et al. (1968) | 1,200± | 69″+ | 34 | 3 | Pentridge Prison |
| Hunter (1968) | 1,021 | 90 percentile + height | 29 | 3 | Approved school |
| Nielsen (1968) | 155 | 180+ cm. | 42 | 2 | Criminal psychopaths |

Table 1—(*Continued*)

| Author(s) and year | Population size | Selection criteria | Number karyotyped | Number 47,XYY | Institution |
|---|---|---|---|---|---|
| | | Prisons and criminally insane | | | |
| Bartlett, Hurley, Brand, & Poole (1968) | 204 | Consecutive admissions | 204 | 2 | Grendon Prison |
| Nielsen, Tsuboi, Tuver, Jensen, & Sachs (1969) | 217 | ? | 97 | 3 | Horsens Prison |
| Marinello et al. (1969) | 2,100 | 183+ cm.; Volunteers | 86 | 2 | Attica State Prison |
| Melnyk, Derencsenyi, Vanasek, Rucci, & Thompson (1969) | 1,400 | 6′+ | 200 | 9 | Atascadero State Hospital |
| Clark et al. (1970) | 2,431 | 71″+ | 124 | 4 | Penal institution |
| | | 71″− | 76 | | |
| | 1,169 | 71″+ | 80 | 3 | Criminally insane |
| | | 71″− | 23 | | |
| | 462 | 71″+ | 60 | 0 | Defective delinquents |
| | | 71″− | 22 | | |
| Court Brown & Smith (1969) | 607 | New entrants—all heights | 607 | 1 | Scottish Borstals |
| | 340 | | 340 | 2 | Approved schools |
| Anonymous—Anatomy of Violence (1970) | 300 | Convicts with abnormal fingerprints | 84 | 0 | Federal penitentiary |
| Falek, Craddick, & Collum (1970) | 2,200 | 6′+; Recurrent criminality; 70 < IQ < 80; or IQ > 80 and psychopathic MMPI. | 22 | 0 | Atlanta Federal Penitentiary |
| Hook & Kim (1970) | 342 | 6′+; Facial acne | 3 | 1 | |
| | | None | 337 | 4 | Nonretarded young offenders |
| Daly (1969a) | ? | 6′+ | 210 | 10 | Ionic State, Norman Beatty Memorial, Central State, and Lima State Hospitals |
| Abdullah, Jarvik, Kato, Johnston, ton, & Lanzkron (1969) | ? | 6′+; Violent, destructive behavior | 49 | 1 | Mental hospitals, hospital for the criminally insane, prison |
| Marinello et al. (1969) | ? | ? | 57 | 1 | Juvenile offenders |
| Spencer, Eyles, & Mason (1969) | ? | 6′+; Criminal tendencies | 5 | 1 | ? |
| Knox & Nevin (1969) | ? | 177.5+ cm. | 67 | 0 | Belfast Prison |
| Marcus & Richmond (1970) | ? | 6′+ | 51 | 2 | Oakalla Prison Farm |
| Davis, McGee, Empson, & Engel (1970) | ? | Long sentences | 190 | 0 | Tennessee State Prison |
| Griffiths, Richards, Zaremba, Abramowicz, & Stewart (1970) | ? | 6′+ | 34 | 2 | Wandsworth Prison |
| | | 5′11″+; Randomly selected | 31 | 1 | |
| | | 5′11″+; Consecutive admissions | 290 | 6 | |
| Anonymous—Anatomy of Violence (1970) | ? | ? | 83 | 6 | Bridgewater |

[a] Male and female.
[b] Personal communication from P. G. McGrath.
[c] Personal communication from M. D. Casey.
[d] Personal communication from J. L. Skinner.
[e] Court Brown, 1968.

TABLE 2

Population Rates of 47,XYY

| Author(s) and year | Sample size | Selection procedure | Number 47,XYY |
|---|---|---|---|
| Sergovich, Valentine, Chen, Kinch, & Smout (1969) | 1,066 | Consecutive live-born males | 4 |
| Lubs & Ruddle (1970) | 2,184 | Consecutively born males | 3 |
| Ratcliffe, Stewart, Melville, Jacobs, & Keay (1970) | 3,496 | Consecutive live-born males | 5 |
| Turner[a] | 1,023 | ? | 3 |
| Jacobs, Brunton, Melville, Brittain, & McClemont (1965) | 266 | Randomly selected male babies | 0 |
| Walzer, Breau, & Gerald (1969) | 1,332 | Phenotypically normal male births | 0 |
| Court Brown, Buckton, Jacobs, Tough, Kuenssberg, & Knox (1966) | 207 | Randomly selected adult males | 0 |
| Recherche Coopérative sur Programme N° 85 (1968) | 6,340 | Males who resembled the general population | 6 |
| | 247 | Normal males | 0 |
| | 896 | Brother, fathers, cousins of ill | 1 |
| | 732 | Maladie neoplasique | 2 |
| Kjessler (1966) | 135 | Fertility clinic ($N = 178$) | 0 |
| Mclree, Price, Court Brown, Tulloch, Newsam, & Maclean (1966) | 143 | Seen at subfertility clinic | 0 |
| | 50 | With low sperm count | 0 |
| Jacobs et al. (1965) | 1,500± | Karyotyped for a variety of reasons | 1 |
| Wilton & Lever (1967) | 750 | Karyotyped for a variety of reasons | 1 |
| Goodman, Miller, & North (1968) | 36 | College basketball players | 0 |
| Price & Jacobs (1970) | 629 | Atomic energy workers above 180 cm. | 0 |

[a] Personal communication from D. S. Borgaonkar, June 16, 1969.

Among the mental hospitals, five studies represent 4,288 patients and found seven 47,XYY subjects among the 377 karyotyped. Two studies representing an unspecified population size karyotyped 86 patients and located 2 more 47,XYY subjects. This rate of 10 among about 8,500 mentally subnormal and mental hospital male patients is distinctly lower than the Carstairs 9 among 342. Probably some 47,XYYs were missed due to screening on the basis of height, and, judging from the Carstairs data, as many as half could have been overlooked. Even so, the discrepancy in rates must cause concern about the possible uniqueness of the Carstairs observation.

Interest in the "aggressive XYY" led to searches in prisons. Sixteen studies, representing approximately 17,500 inmates, found 55 cases of 47,XYY in the 2,663 karyotyped. This would seem to be higher than the mentally retarded populations, and somewhat higher than the mental hospital populations. In some of these studies there was no selection for height, and that methodological difference may be important for the observed prevalence figure. Eight of these studies on a total of 3,803 detained individuals account for 36 of the 47,XYY cases; the other 10,744 inmates account for only 18 of the 47,XYY cases. These rates are, however, much below the Carstairs findings. Eight other studies, representing populations of unspecified size, reported 29 more 47,XYY cases among the 1,018 karyotyped; the studies cannot help establish prevalence rates without some knowledge of their respective population sizes. Also, it may be an error to add frequencies among these studies, for there may be actual between-group differences that should not be ignored.

Although these summary prevalence figures are a good bit less impressive than the Carstairs rate, the mental hospital data and the prison data indicate rates four or five times the *presumed* general population rate. Prior to an empirical estimate of the population incidence, the presumed rate of 1/2000 males was arrived at by considering that XYY would result from nondisjunction during miosis. Klinefelter's syndrome (XXY) also resulted from nondisjunction, and its prevalence was known. Klinefelter's could result from either maternal or paternal nondisjunction, but the XYY could result only from paternal nondisjunction. Hence the estimate was half the XXY prevalence. This initially presumed rate for XYY now seems too low.

Interest in the 47,XYY genotype led to some attempts to ascertain empirically the incidence in the population, despite the very considerable expense involved. Part of the studies contained in Table 2 bear upon this issue.

Four studies on consecutively born males represent 7,769 births and reported 15 cases of 47,XYY—a rate four times that previously expected! One other report, however, on 1,332 newborns failed to locate a single 47,XYY case; this study dealt only with phenotypically normal males, and did not deal with consecutive births, a methodological distinction that may contribute to the discrepancy.

The composite rate among consecutive newborns is 1 in 518 births, a rate not very different from that found in the institutions. Whatever support the "aggressive XYY" hypothesis might receive from a differential rate among prisoners will not be confirmed by this rate of incidence among the newborns—unless they all grow up with a history of imprisonment.

TABLE 3

PARENTAL AGE AT BIRTH OF 47,XYY OFFSPRING

| Age | Maternal percentage | Maternal controls[a] | Paternal percentage |
|---|---|---|---|
| 15–19 | 9.46 | 4.06 | 1.43 |
| 20–24 | 27.03 | 25.47 | 14.29 |
| 25–29 | 23.68 | 31.54 | 22.86 |
| 30–34 | 21.62 | 22.42 | 28.57 |
| 35–39 | 9.46 | 12.42 | 18.57 |
| 40–44 | 4.05 | 3.78 | 7.14 |
| 45–49 | 2.70 | .30 | 5.71 |
| 50–54 | 1.35 | — | |
| | | | 1.43 |

Note.—Age is in years. Percent signs are omitted. This table is based on reports of Abdullah et al., 1969; Balodimos, Lisco, Irwin, Merrill, & Dingman, 1966; Bartlett, Hurley, Brand, & Poole, 1968; Borgaonkar, Murdoch, McKusick, Borkowf, Money, & Robinson, 1968; Court Brown, Harnden, Jacobs, Maclean, & Mantle, 1964; Court Brown, Price, & Jacobs, 1968a; Cowie & Kahn, 1968; Daly, Chun, Ewanowski, & Osborne, 1969; Dunn, Ford, Auersperg, & Miller, 1961; Fraccaro, Glen Bott, Davies, & Schutt, 1962; Griffiths, Richards, Zaremba, Abramowicz, & Stewart, 1970; Hashi, Tsutsumi, & Tsuda, 1969; Hauschka, Hasson, Goldstein, Koepf, & Sandberg, 1962; Hayward & Bower, 1960; Hustinx & van Olphen, 1963; Kosenow & Pfeiffer, 1966; Marcus & Richmond, 1970; Marinello, Berkson, Edwards, & Bannerman, 1969; Noel, Quack, Durand, & Rethoré, 1969; Persson, 1967; Price, Strong, Whatmore, & McClemont, 1966; Ratcliffe, Stewart, Melville, Jacobs, & Keay, 1970; Ricci & Malacarne, 1964; Richards & Stewart, 1966; Sandberg, Ishihara, Crosswhite, & Koepf, 1963; Sergovich, Valentine, Chen, Kinch, & Smout, 1969; Spencer, Eyles, & Mason, 1969; Thorburn, Chutkan, Richards, & Bell, 1968; Tzoneva-Maneva, Bosajleva, & Petrov, 1966; Vigi, Volpato, Zardini, Gaburro, & Tiepolo, 1969; Wiener & Sutherland, 1968; Wiener, Sutherland, & Bartholomew, 1968.

[a] Penrose (1966).

The difference in prevalence of 47,XYY and 47,XXY may result from different viability. In a study of spontaneous abortions, 14 sex chromosome triploids were found (Polani, 1966). Only 1 of these was a 47,XYY subject—9 were 47,XXY, and 4 were 47,XXX. One might speculate on Polani's data and their implications for the "surprisingly" high incidence among the newborns of 47,XYY compared with 47,XXY. Perhaps, in fact, 47,XYY is the most viable of the sex chromosome triploids.

### PARENTAL AGE AND BIRTH ORDER

Parental age at birth for autosomal aneuploidies has been studied since Jenkin's (1933) report on Down's syndrome. Only recently has it become possible for this interest to extend to other aneuploid states. In the case of 47,XYY, parental ages at birth of the offspring are provided in a number of cases. The distribution of maternal age by 5-year intervals is given in Table 3 for 74 cases. The mean maternal age of 28.69 years with a standard deviation of 7.42 is distinctly not significantly different from a control population mean of 28.35 with a standard deviation of 5.95 (Penrose, 1966). Chi-square on the distributions requires that ages 15–24 years be combined, and that ages 35–49 years be combined, to provide expected frequencies of five per cell; when this is done, the $\chi^2$ of 2.48 is not significant ($.5 > p > .3$, $df = 3$). However, this is not a very powerful test. If expected frequencies of only three are accepted, then a $\chi^2$ of 8.55 ($df = 5$) slightly suggests a deviation ($.25 > p > .10$), with the major contributions due to both extremes of the age range.

Paternal age on 70 cases also is noticeably not different from a control group (Penrose, 1955). The 47,XYY mean paternal age is 31.84 years with a standard deviation of 7.30; the control paternal age mean is 30.90. The distribution of paternal ages by 5-year intervals is provided in Table 3.

The age difference between parents (paternal minus maternal) is greater than in Penrose's (1955) control group. The data on 68 cases are shown in Table 4. This mean

TABLE 4

Paternal Minus Maternal Age Distribution

| Age difference | −5 | −4 | −3 | −2 | −1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 1 | 1 | 2 | 1 | 2 | 5 | 9 | 10 | 7 | 4 | 6 | 5 | 7 | 2 | 6 |

difference is 3.62 years, compared with the controls' of 2.30 years. However, three of the 47,XYY have differences of 14, 14, and 17 years; without them, the mean difference would be only 3.11, suggesting that the extreme cases may be unduly influencing the estimate of central tendency. Of the 68 cases, 31 are below the control mean difference, and 37 are above it ($\chi^2 = .53$, $.50 > p > .25$, $df = 1$). If there is an effect with paternal age (with maternal age held constant), it is not strong enough to be confirmed by these data.

Birth rank of the 47,XYY males is given in 50 cases (Table 5). The range is from 1 to 16, with a median of 2 and a mode of 1. The number of siblings in the family is provided in 33 cases, where the range is from 1 to 16, with a median of 4 and a mode of 3. The distributions are provided in Table 6.

## MEDICAL DATA ON 47,XYY MALES

Bias introduced by the case study method deserves careful consideration. Nevertheless, it would seem imprudent simply to ignore the cases already known to possess a 47,XYY complement. Disorders of sexual development and fertility, bone deformities, dental irregularities, cardiac symptomatology, dermatoglyphics, and hormone levels are among the conditions reported in cases found to have a 47,XYY karyotype.

TABLE 5

Birth Rank Distribution

| Birth rank | 1 | 2 | 3 | 4 | 5 | 6+ |
|---|---|---|---|---|---|---|
| Frequency | 17 | 11 | 8 | 3 | 4 | 7 |

Note.—This table is based on reports of Abdullah, Jarvik, Kato, Johnston, & Lanzkron, 1969; Balodimos, Lisco, Irwin, Merrill, & Dingman, 1966; Bartlett, Hurley, Brand, & Poole, 1968; Borgaonkar, Murdoch, McKusick, Borkowf, Money, & Robinson, 1968; Boucharlat & Jalbert, 1969; Court Brown, Harnden, Jacobs, Maclean, & Mantle, 1964; Cowie & Kahn, 1968; Daly, 1969a; Daly, Chun, Ewanowski, & Osborne, 1969; Dunn, Ford, Auersperg, & Miller, 1961; Elbualy, 1969; Fraccaro, Glen Bott, Davies, & Schutt, 1962; Hashi, Tsutsumi, & Tsuda, 1969; Hustinx & van Olphen, 1963; Leff & Scott, 1968; Marcus & Richmond, 1970; Mirouze, Emberger, Jaffiol, Orsetti, & Meynadier, 1966; Parker, Melnyk, & Fish, 1969; Parker, Mavalwala, Weise, Koch, Hatashita, & Cibilich, 1970; Persson, 1967; Price, Strong, & Whatmore, 1966; Richards & Stewart, 1966; Sandberg, Ishihara, Crosswhite, & Koepf, 1963; Sandberg, Koepf, Ishihara, & Hauschka, 1961; Sergovich, Valentine, Chen, Kinch, & Smout, 1969; Spencer, Eyles, & Mason, 1969; Thorburn, Chutkan, Richards, & Bell, 1968; Vigi, Volpato, Zardini, Gaburro, & Tiepolo, 1969; Wiener, Sutherland, Bartholomew, & Hudson, 1968.

### *Disorders of Sexual Development*

Genital abnormality is among the earliest suspected sequelae of the 47,XYY karyotype. Among the 10 case reports by 1965 that mentioned the condition of the genitalia, 8 cited abnormalities. However, beginning with 1966, another 85 cases appeared, 13 citing anomalies and 72 noting normal genitalia. Among the 13 anomalies were low sperm count in two cases (Marcus & Richmond, 1970; Wilton & Lever, 1967), a small right testis following sterilization (Persson, 1967), a monozygotic twin with an undescended testicle (the co-twin presumably was normal; Rainer, Jarvik, Abdullah, & Kato, 1969), and a circumcision at age 3 because of phimosis (Elbualy, 1969). Among the eight cases prior to 1966, one of those anomalies was undescended testis—however, his two 46,XY brothers also had undescended testes. Other sexual anomalies reported are shown in Table 7, as are reports of normal genitalia.

The possible transmission of the 47,XYY genome has been considered since the father and son pair of 47,XYY reported by Tzoneva-Maneva, Bosajleva, & Petrov (1966). Only one other father–son pair has been reported (Sundequist & Hellström, 1969). In fact if such transmission does occur, it would seem to be unusual. Thompson, Melnyk, and Hecht (1967) first reported only normal 46,XY spermatogonial metaphases in a 47,XYY male. Similar observations have been reported

TABLE 6

Sibship Size Distribution

| Total sibs in family | 1 | 2 | 3 | 4 | 5 | 6+ |
|---|---|---|---|---|---|---|
| Frequency | 3 | 4 | 9 | 5 | 4 | 8 |

TABLE 7

SEXUAL ANOMALIES

| Description | Author(s) and year |
|---|---|
| Hypogonadism | Balodimos, Lisco, Irwin, Merrill, & Dingman, 1966; Buckton, Bond, & McBridge, 1962; Court Brown, Harnden, Jacobs, Maclean, & Mantle, 1964; Lehrnbecher & Lucas, 1969a; Mirouze, Emberger, Jaffiol, Orsetti, & Meynadier, 1966; Montero & Duran, 1965; Thorburn, Chutkan, Richards, & Bell, 1968. |
| Small testes | Milcu, Negoescu, Maximilian, Garoiu, Augustin, & Iliescu, 1964; Thorburn et al., 1968; Wilton & Lever, 1967 |
| Small scrotum | Court Brown et al., 1964 |
| Undescended testes | Dunn, Ford, Auersperg, & Miller, 1961; Fraccaro, Glen Bott, Davies, & Schutt, 1962; Kosenow & Pfeiffer, 1966; Pergament, Sato, Berlow, & Mintzer, 1968; Rainer, Jarvik, Abdullah, & Kato, 1969; Sandberg, Ishihara, Crosswhite, & Koepf, 1963 |
| Late descending testis | Fraccaro et al., 1962 |
| Small penis | Court Brown et al., 1964; Dunn et al., 1961; Milcu et al., 1964; Mirouze et al., 1966 |
| Gynecomastia | Balodimos et al., 1966; Parker, Melnyk, & Fish, 1969; Parker, Mavalwala, Weise, Koch, Hatashita, & Cibilich, 1970 |
| Scanty pubic hair | Court Brown et al., 1964; Marinello, Berkson, Edwards, & Bannerman, 1969; Mirouze et al., 1966 |
| Low sperm count | Hultén, 1970; Marcus & Richmond, 1970; Tettenborn, Gropp, Murken, Tinnefeld, Fuhrmann, & Schwinger, 1970; Wilton & Lever, 1967 |
| Other anomaly | Elbualy, 1969; Hultén, 1970; Nielsen, Christensen, Johansen, & Frøland, 1966; Persson, 1967; Skakkeback, 1970; Tettenborn et al., 1970 |
| Genitalia normal | Abdullah, Jarvik, Kato, Johnston, & Lanzkron, 1969; Chaptal, Jean, Bonnet, Emberger, Meylan, & Riev, 1968; Clark, Telfer, Baker, & Rosen, 1970; Cleveland, Arias, & Smith, 1969; Court Brown, 1968; Court Brown, Price, & Jacobs, 1968a; Cowling, Rigo, & Martin, 1969; Hustinx & van Olphen, 1963; Kelly, Almy, & Bernard, 1967; Leff & Scott, 1968; Marinello et al., 1969; Noel, Quack, Durand, & Rethoré, 1969; Papanicolaou, Kirkham, & Loraine, 1968; Price, Strong, Whatmore, & McClemont, 1966; Rainer et al., 1969; Ratcliffe, Stewart, Melville, Jacobs, & Keay, 1970; Ricci & Malacarne, 1964; Richards & Stewart, 1966; Sergovich, Valentine, Chen, Kinch, & Smout, 1969; Spencer, Eyles, & Mason, 1969; Vigi, Volpato, Zardini, Gaburro, & Tiepolo, 1969; Wiener & Sutherland, 1968; Wiener, Sutherland, Bartholomew, & Hudson, 1968. |

in 18 more 47,XYY males, most of them examined for mosaicism (Evans, Ford, Chaganti, Blank, & Hunter, 1970; Hsu, Shapiro, & Hirschhorn, 1970; Hultén, 1970; Melnyk, Thompson, Rucci, Vanasek, & Hayes, 1969; Sundequist & Hellström, 1969; Tettenborn, Schwinger, & Gropp, 1970). In a few of the cases some XY/XYY mosaic spermatogonia were found. More recently, however, a reexamination for fluorescence (F body) of Hultén's (1970) case indicated that 45% of the primary spermatocytes contained two Y chromosomes and 2.95% of the spermatozoa contained two Y chromosomes. Three control subjects had two Y chromosomes in from 1.24% to 1.70% of their spermatozoa (Hultén & Pearson, 1971). Diaso and Glass (1970) reported another case of 47,XYY to have an F body in 70% of the spermatozoa and two F bodies in another 5%. For comparison, Sumner, Robinson, and Evans (1971) have reported two F bodies in 1.26% of the spermatozoa of nine 46,XY males.

Children of 47,XYY fathers are present in at least 10 cases (Daly, 1969a; Falek, Craddick, & Collum, 1970; Lisker, Zenzes, & Fonesca, 1968; Marcus & Richmond, 1970; Sandberg et al., 1961; Stenchever & Macintyre, 1969; Wiener, Sutherland, Bartholomew, & Hudson, 1968; Wilton & Lever, 1967). Poor reproductive histories appear in three cases (Sandberg et al., 1961; Stenchever & Macintyre, 1969; Wilton & Lever, 1967). Absence of offspring is specifically noted in seven instances (Bartlett, Hurley, Brand, & Poole, 1968; Goodman, Smith, & Migeon, 1967; Parker, Melnyk, & Fish, 1969; Persson, 1967; Wiener et al., 1968). Most of the reported children of 47,XYY fathers have been 46,XX and 46,XY.

Three of the cases report 47,XYY chromosome complements associated with female external genitalia (Boczkowski, 1969, 1970; Franks, Bunting, & Engel, 1967; Vignetti, Capotori, & Ferrante, 1964). This seriously raises the question of mosaicism, with at least one of these suspected of being 45,X/47,XYY mosaic. In another, both peripheral blood and skin cultures were made, both being 47,XYY. Still the possibility remains of all three being mosaic.

### *Disorders of Bones and/or Joints*

Tzoneva-Maneva et al. (1966) suggested that XYY led to idiopathic osteoarthropathy since they found the condition in an XYY boy and his XYY father. Their interpretation of the data is erroneous (see, e.g., Rimoin & Borgaonkar, 1966), but various bone and joint anomalies have been reported. These cases are documented in Table 8. Normal bone age has also been cited. Reports on 41 cases specifically mentioned presence or absence of some bone anomaly. Selection bias would seem to be of paramount importance here, as only two of the studies involved surveys; in all the others the anomaly presumably influenced the decision to seek a karyotype.

### *Dental Irregularities*

Dental irregularities have been cited in a few of the case reports. Enamel hypoplasia, teeth in irregular order, unerupted teeth, malocclusion, retained deciduous upper cuspids in a 41-year-old, and chalky teeth have all been described (Clark, Telfer, Baker, & Rosen, 1970; Cowling, Rigo, & Martin, 1969; Dallapiccola & Malacarne, 1970; Dunn, Ford, Auersperg, & Miller, 1961; Kosenow & Pfeiffer, 1966; Ricci & Malacarne, 1964; Wiener & Sutherland, 1968). Teeth in good condition has been cited also (Parker et al., 1969). These cases account for only 13 of the known 47,XYY males. They would thus not seem to be strong evidence for the presence of a particular tendency toward dental irregularities among the 47,XYY males as a group.

TABLE 8

BONE AND/OR JOINT ANOMALIES

| Description | Author(s) and year |
|---|---|
| Face—protruding mandible, large jaw, maxillary overbite, mandibulofacial dysostosis, slightly short lower jaw | Abdullah, Jarvik, Kato, Johnston, & Lanzkron, 1969; Castilla, Breg, & Miller, 1964; Cowling, Rigo, & Martin, 1969; Dunn, Ford, Auersperg, & Miller, 1961; Kosenow & Pfeiffer, 1966; Parker, Melnyk, & Fish, 1969; Parker, Mavalwala, Weise, Koch, Hatashita, & Cibilich, 1970 |
| Head—long narrow head, elongated skull, receding forehead, skull anomalies | Dallapiccola & Malacarne, 1970; Daly, 1969b; Daly, Chun, Ewanowski, & Osborne, 1969; Hustinx & van Olphen, 1963; Ricci & Malacarne, 1964; Sergovich, Valentine, Chen, Kinch, & Smout, 1969; Spencer, Eyles, & Mason, 1969 |
| Hands and feet—brachyphalangia, flexion deformities in fingers, short incurved fifth fingers, talipes equinovares, outsize big toe | Buckton, Bond, & McBride, 1962; Court Brown, Harnden, Jacobs, Maclean, & Mantle, 1964; Dallapiccola & Malacarne, 1970; Hustinx & van Olphen, 1963; Spencer et al., 1969; Wiener & Sutherland, 1968 |
| Arms and legs—bilateral genu valgum, bilateral genu varum, deformed knees and elbows, radioulnar synostosis | Cleveland, Arias, & Smith, 1969; Court Brown, Price, & Jacobs, 1968b; Elbualy, 1969; Lehrnbecher & Lucas, 1969a; Mirouze, Emberger, Jaffiol, Orsetti, & Meynadier, 1966; Parker et al., 1969; Thorburn, Chutkan, Richards, & Bell, 1968 |
| Other—spine scoliosis, mildly kyphotic back without scoliosis, elevated shoulders, hip joint that could not be fully abducted, generalized osteoporosis, retarded bone age, idiopathic osteoarthropathy | Bosajleva, 1968; Dallapiccola & Malacarne, 1970; Daly, 1969b; Dunn et al., 1961; Kosenow & Pfeiffer, 1966; Parker et al., 1969; Pergament, Sato, Berlow, & Mintzer, 1968; Persson, 1967; Tzoneva-Maneva, Bosajleva, & Petrov, 1966 |
| Normal bone age | Daly et al., 1969; Parker et al., 1970; Ricci & Malacarne, 1964; Vigi, Volpato, Zardini, Gaburro, & Tiepolo, 1969 |

Note.—Citation for a group of statements implies that the report contained at least one of the terms in the group.

### Skin Disorders

Acne is another condition that has been associated with the XYY in the popular press. The case reports note its presence in varying degree 16 times (Abdullah, Jarvik, Kato, Johnston, & Lanzkron, 1969; Clark et al., 1970; Falek et al., 1970; Marinello, Berkson, Edwards, & Bannerman, 1969; Sandberg et al., 1961; Santen, DeKrester, Paulsen, & Vorhees, 1970; Vigi, Volpato, Zardini, Gaburro, & Tiepolo, 1969); 7 of the 16 were found by screening at a dermatology service. In four cases, acne specifically was not observed (Bartholomew & Sutherland, 1969; Clark et al., 1970), nor was there any old scarring from acne. Although adult XYY males may have an unnoted history of acne, considering the number of case reports with reasonably thorough medical examinations, this low prevalence would not seem to support the notion that 47,XYY men are particularly prone to develop skin disorders.

### Cardiac Disorders

Price (1968) compared 18 men with 47,XYY chromosomes to a group of 28 controls with 46,XY constitutions and found an abnormally high P–R interval in the electrocardiogram (EKG) of six of the XYY and none of the XY males. Interest focused for a time on possible cardiac symptomatology associated with the 47,XYY complement. Reports on 36 men with 47,XYY mention cardiac symptoms or their absence. These data are shown in Table 9. They fail to reveal any very specific cardiac anomaly, and the decision to karyotype those cases found to be 47,XYY (15 of them) was influenced to an indeterminate degree by the presence of cardiac symptomatology. However, examination of the P–R interval in 47,XYY patients identified by mass screening programs would seem appropriate.

### Neurological Disorders

Neurological abnormalities have been suggested to result from 47,XYY constitution (Daly, 1969b; Forssman & Hambert, 1966, 1967). Detection of mild or diffuse involvement might prove particularly difficult. Testing procedures for minimal dysfunctions must be sensitive to very subtle deficits in behavior. It would seem appropriate to examine subjects with 47,XYY karyotype identified in mass screenings, and to examine them in sufficient neuropsychological detail to dis-

TABLE 9

CARDIAC ANOMALIES

| Description | Author(s) and year |
|---|---|
| Abnormal EKG | Hustinx & van Olphen, 1963; Lehrnbecher & Lucas, 1969a |
| Borderline EKG | Marcus & Richmond, 1970 |
| Normal EKG | Cowling, Rigo, & Martin, 1969; Daly, Chun, Ewanowski, & Osborne, 1969; Mirouze, Emberger, Jaffiol, Orsetti, & Meynadier, 1966; Parker, Melnyk, & Fish, 1969; Parker, Mavalwala, Weise, Koch, Hatashita, & Cibilich, 1970; Thorburn, Chutkan, Richards, & Bell, 1968; Wiener & Sutherland, 1968 |
| P-R interval lengthened | Borgaonkar, Murdoch, McKusick, Borkowf, Money, & Robinson, 1968 |
| P-R interval normal | Borgaonkar & Mules, 1970; Borgaonkar et al., 1968; Wiener & Sutherland, 1968 |
| Bradycardia | Richards & Stewart, 1966 |
| Normal pulse | Price, 1968 |
| Moderately high blood pressure | Marinello, Berkson, Edwards, & Bannerman, 1969 |
| Normal blood pressure | Cowling et al., 1969; Hustinx & van Olphen, 1963; Richards & Stewart, 1966 |
| Diastolic murmur | Price, 1968 |
| Systolic murmur | Hustinx & van Olphen, 1963 |
| Apical systolic bruit | Marinello et al., 1969 |
| Slight heart enlargement | Hustinx & van Olphen, 1963 |
| Normal sized heart | Daly et al., 1969; Price, 1968 |
| Congenital heart disease | Emerit, de Grouchy, Vernant, & Corone, 1967; Price, 1968 |
| Myocardial infarct | Balodimos, Lisco, Irwin, Merrill, & Dingman, 1966; Dallapiccola & Malacarne, 1970 |

TABLE 10

NEUROLOGICAL ANOMALIES

| Description | Author(s) and year |
|---|---|
| Grossly abnormal EEG | Forssman, 1967 |
| Abnormal EEG | Boucharlat & Jalbert, 1969; Noel, Quack, Durand, & Rethoré, 1969; Parker, Melnyk, & Fish, 1969; Pergament, Sato, Berlow, & Mintzer, 1968; Wiener, Sutherland, Bartholomew, & Hudson, 1968 |
| Spike focus in EEG | Borgaonkar, Murdoch, McKusick, Borkowf, Money, & Robinson, 1968; Elbualy, 1969; Lehrnbecher & Lucas, 1969a |
| Abnormal paroxysmal EEG | Marcus & Richmond, 1970 |
| Generalized moderate cerebral dysrhythmia | Parker et al., 1969 |
| Mildly abnormal EEG | Bartlett, Hurley, Brand, & Poole, 1968; Borgaonkar et al., 1968; Cowie & Kahn, 1968; Daly, 1969a; Welch, Borgaonkar, & Herr, 1967; Wiener & Sutherland, 1968 |
| Normal EEG | Borgaonkar et al., 1968; Chaptal, Jean, Bonnet, Emberger, Meylan, & Riev, 1968; Clark, Telfer, Baker, & Rosen, 1970; Cleveland, Arias, & Smith, 1969; Daly, 1969a; Daly, Chun, Ewanowski, & Osborne, 1969; Dunn, Ford, Auersperg, & Miller, 1961; Anonymous—An English XYY Murder Trial, 1969; Falek, Craddick, & Collum, 1970; Leff & Scott, 1968; Mirouze, Emberger, Jaffiol, Orsetti, & Meynadier, 1966; Nielsen & Pedersen, 1969; Noel, Quack, Durand, & Rethoré, 1969; Parker, Mavalwala, Weise, Koch, Hatashita, & Cibilich, 1970; Persson, 1967; Spencer, Eyles, & Mason, 1969 |
| Skull films normal | Elbualy, 1969; Marcus & Richmond, 1970 |
| Neurological exam normal | Clark et al., 1970; Parker et al., 1969; Parker et al., 1970; Price, 1968; Sandberg, Ishihara, Crosswhite, & Koepf, 1963 |
| Cerebellar tremor | Daly, 1969b |
| Progressive neurological deficit | Daly, 1969a; Lehrnbecher & Lucas, 1969a |
| History of epilepsy or seizures | Boucharlat & Jalbert, 1969; Daly, 1969b; Forssman, 1967; Fraccaro, Glen Bott, Davies, & Schutt, 1962; Hayward & Bower, 1960; Kelly, Almy, & Bernard, 1967; Noel, Quack, Durand, & Rethoré, 1969; Rainer, Jarvik, Abdullah, & Kato, 1969; Welch et al., 1967 |
| Brief losses of consciousness | Vigi, Volpato, Zardini, Gaburro, & Tiepolo, 1969 |
| No history of seizures | Elbualy, 1969; Nielsen & Pedersen, 1969; Persson, 1967; Vigi et al., 1969 |

close minimal behavioral deficit should it be present.

The records of 79 patients contain information relevant to neurological status. These data are summarized in Table 10. Some degree of abnormality is cited in 24 of the 51 cases referred for electroencephalogram (EEG). Although these data are hardly less subject to sampling biases than are the data on other phenotypic attributes discussed, the hypothesis of minor neurological involvement remains an interesting one. None of the case reports yet published reflects the intensity of study that would seem appropriate.

### *Hormone Levels*

Testosterone levels were higher than normal in six 47,XYY Carstairs inmates ($p < .02$), but not higher than matched hospital controls (Ismail, Harkness, Kirkham, Loraine, Whatmore, & Brittain, 1968). Nine 47,XYY males from Rampton hospital also had higher than normal levels, but not higher than matched hospital controls (Rudd, Galal, & Casey, 1968). A group of 12 institutionalized 47,XYY (10 mental subnormals, 1 mentally ill, and 1 epileptic) had levels insignificantly different from controls and from ambulatory normals (Price & van der Molen, 1970). Nine other 47,XYY males had testosterone levels below their controls (Plasse, Patricot, Giboulet, Noel, & Revol, 1970). One case report shows testosterone level four standard deviations above normal (Welch, Borgaonkar, & Herr, 1967), but 17 other cases are reported to have levels within the normal range or within a suitably matched control population range (Boucharlat & Jalbert, 1969;

Goodman et al., 1967; Hudson, Burger, Wiener, Sutherland, & Bartholomew, 1969; Parker et al., 1969; Santen et al., 1970; Shapiro, 1970b). Wide fluctuations in testosterone levels have been observed among the studies and within patients, and differences between laboratories make comparison among the studies difficult.

Pituitary gonadotropins have been reported in four cases; one case had a high level (Persson, 1967), but the other three were normal (Balodimos, Lisco, Irwin, Merrill, & Dingman, 1966; Cowling et al., 1969; Parker et al., 1969). Follicle-stimulating hormone (FSH) was high in one case (Shapiro, 1970a). The FSH of another patient was high in a group of seven patients; the other six were within the control range (Santen et al., 1970). Eleven more 47,XYY males were reported to have FSH levels within the normal range (Goodman et al., 1967; Lundberg & Wahlstrom, 1970; Marcus & Richmond, 1970; Mirouze, Emberger, Jaffiol, Orsetti, & Meynadier, 1966; Nielsen, Christensen, Johansen, & Frøland, 1966; Papanicolaou, Kirkham, & Loraine, 1968). Serum luteinizing hormone (LH) was reported to be grossly elevated in three 47,XYY males, differing from their controls at the .01 level (Papanicolaou et al., 1968). The LH in another group of seven XYY men was higher than their control group ($p < .001$; Parker, 1969). Another case report was also of high LH, but again two of five controls were higher (Shapiro, 1970b). One 47,XYY from a group of seven had high LH, but the other six were within the control group range (Santen et al., 1970). One uncontrolled case report was high for LH (Parker, Mavalwala, Weise, Koch, Hatashita, & Cibilich, 1970), but 10 other cases were within the normal or control range (Goodman et al., 1967; Hudson et al., 1969; Lundberg & Wahlstrom, 1970). Again, methodological problems in measuring these levels may justify hesitancy in interpreting these pituitary gonadotropin levels.

Growth hormone has been assessed in nine 47,XYY cases—all adults, but one who was aged 16. No significant abnormalities have been reported (Lundberg & Wahlstrom, 1970; Nielsen, Johansen, & Yde, 1969; Nielsen, Yde, & Johansen, 1969, 1970). Growth hormone levels need to be determined longitudinally on such cases as may be identified at birth.

In 27 cases 17-ketosteroids have been examined. No significant deviations have been reported (Boucharlat & Jalbert, 1969; Cleveland, Arias, & Smith, 1969; Cowling et al., 1969; Daly, Chun, Ewanowski, & Osborne, 1969; Lehrnbecher & Lucas, 1969a; Marcus & Richmond, 1970; Nielsen, Tsuboi, Tuver, Jensen, & Sachs, 1969; Nielsen, Yde, & Johansen, 1969; Parker et al., 1969; Persson, 1967; Plasse et al., 1970; Spencer, Eyles, & Mason, 1969). In six cases 17-hydroxycorticosteroids have been measured; again no significant abnormalities have been cited (Boucharlat & Jalbert, 1969; Cowling et al., 1969; Daly et al., 1969; Lehrnbecher & Lucas, 1969a; Persson, 1967). Peroxidase values of neutrophil polymorphonuclears in 47,XYY males were lower than ambulatory controls (Noel, Quack, & Talon, 1969).

### *Dermatoglyphic Alterations*

Dermatoglyphic examinations have been reported in 42 cases. It has been suggested that total finger tip ridge count (TRC) is lower in 47,XYY males than in normals. In the 20 cases with a TRC given, the range is from 49 to 244; 10 of these cases fall above the normal mean of 145 (Penrose, 1967), and 10 of them fall below it. In fact, the median TRC is 150—above the normal mean (Bartlett et al., 1968; Borgaonkar, Murdoch, McKusick, Borkowf, Money, & Robinson, 1968; Cleveland et al., 1969; Marcus & Richmond, 1970; Marinello et al., 1969; Parker et al., 1969; Parker et al., 1970; Tsuboi & Nielsen, 1969). Polani (1969) reported data relating to several sex chromosome aneuploid states which, despite the above 20 case reports, remain provocative; and Penrose (1967) formulated an expression for the expected ridge count (ERC) in aneuploid states:

$$\mathrm{ERC} = 187 - 30N_{\mathrm{X}} - 12N_{\mathrm{Y}},$$

where $N_{\mathrm{X}}$ and $N_{\mathrm{Y}}$ represent the number of X and Y chromosomes in the karyotype—an expression that fits his observations on 47,XYY quite well. Borgaonkar and Mules

(1970) reported a mean TRC of 110 for 28 cases of XYY; the mean TRC is significantly lower than for normal males ($p < .001$).

Seven cases give the *a-b* ridge count (ridge count between palmar triradii beneath first and second fingers; Mavalwala, Parker, & Melnyk, 1969; Tsuboi & Nielsen, 1969). They range from 68 to 94 with a median of 83. The normal 46,XY *a-b* ridge count is 87.9 with a standard deviation of 8.6 (Tsuboi & Nielsen, 1969).

The maximal axial triradii (*atd*) angle has been reported 10 times (Bartlett et al., 1968; Daly et al., 1969; Spencer et al., 1969; Tsuboi & Nielsen, 1969), and in two other cases it was noted as normal (Cleveland et al., 1969). The maximal *atd* angle ranges from 34 to 124 in these patients, with a median of 92.5. The normal maximal *atd* angle is 84.7 with a standard deviation of 1.3 (Tsuboi & Nielsen, 1969).

Presence of an arch pattern on fingertips has been reported as increased in some 47,XYY males, occurring in 15.5% compared with only 4.9% of normals ($p < .001$; Tsuboi & Nielsen, 1969). Borgaonkar and Mules (1970) reported a similar finding, with arches present in 17.7% of 27 cases of 47,XYY.

A significantly low incidence of tibial loops in distal hypothenar areas of the soles of the feet has been reported—14.3% in XYY and 25% in controls (Mavalwala, Thompson, Parker, & Wilson, 1970). Simian creases on palms have been described in two cases (Daly et al., 1969; Mavalwala et al., 1970), and specifically noted as absent in six cases (Tsuboi & Nielsen, 1969). Whatever dermatoglyphic differences may actually exist in 47,XYY males, further data must be awaited to delineate their nature.

### *Birth and Early Development*

Birth weights were reported on 21 cases (Borgaonkar et al., 1968; Boucharlat & Jalbert, 1969; Cowie & Kahn, 1968; Daly et al., 1969; Dunn et al., 1961; Elbualy, 1969; Forssman, 1967; Marcus & Richmond, 1970; Parker et al., 1970; Ratcliffe, Stewart, Melville, Jacobs, & Keay, 1970; Sandberg, Ishihara, Crosswhite, & Koepf, 1963; Sergovich, Valentine, Chen, Kinch, & Smout, 1969; Vigi et al., 1969). They range from 2.11 kilograms to 5.23 kilograms; the median is 3.29 kilograms. Three of the cases were breech presentations (Dunn et al., 1961; Parker et al., 1970; Persson, 1967). In six instances, a normal delivery was noted (Cowie & Kahn, 1968; Daly, 1969a; Daly et al., 1969; Marinello et al., 1969; Parker et al., 1969). Eight cases are noted as having walked between 12 and 24 months of age (Abdullah et al., 1969; Boucharlat & Jalbert, 1969; Cowie & Kahn, 1968; Daly et al., 1969; Mirouze et al., 1966; Parker et al., 1970; Sandberg et al., 1963; Spencer et al., 1969).

### *Height, Weight, and Body Build*

Increased stature has been implicated as a concomitant of extra Y chromosomes since the report of Bray and Josephine (1963). Information has been provided on the size of 47,XYY men in several cases (Abdullah et al., 1969; Anonymous—An English XYY Trial, 1969; Balodimos et al., 1966; Bartlett et al., 1968; Borgaonkar et al., 1968; Buckton, Bond, & McBride, 1963; Cleveland et al., 1969; Court Brown, Harnden, Jacobs, Maclean, & Mantle, 1964; Court Brown, Price, & Jacobs, 1968a; Cowie & Kahn, 1968; Cowling et al., 1969; Daly, 1969b; Daly et al., 1969; Dunn et al., 1961; Emerit, de Grouchy, Vernant, & Corone, 1967; Falek et al., 1970; Forssman, 1967; Forssman, Akesson, & Wallin, 1968; Griffiths, Richards, Zaremba, Abramowicz, & Stewart, 1970; Hashi, Tsutsumi, & Tsuda, 1969; Kosenow & Pfeiffer, 1966; Leff & Scott, 1968; Lehrnbecher & Lucas, 1969a; Lisker et al., 1968; Lundberg & Wahlstrom, 1970; Marcus & Richmond, 1970; Marinello et al., 1969; Mirouze et al., 1966; Noel, Quack, Durand, & Rethoré, 1969; Papanicolaou et al., 1968; Parker et al., 1969; Parker et al., 1970; Persson, 1967; Price et al., 1966; Rainer et al., 1969; Ricci & Malacarne, 1964; Richards & Stewart, 1966; Sandberg et al., 1963; Shapiro, 1970b; Spencer et al., 1969; Stenchever & Macintyre, 1969; Telfer, Baker, & Longtin, 1968; Thorburn, Chutkan, Richards, & Bell, 1968; Turner & Jennings, 1961; Vigi et al., 1969; Wiener & Sutherland, 1968; Wiener et al., 1968; Wilton & Lever, 1967). Of the 99 records giving height for patients over 15 years old, 80 are over 180 centimeters tall. The range is from 143 to 227 centimeters, with a median of 185 centimeters. However, most of these patients were selected for karyotyping because of their tall stature. Even among those case reports that were not part of institutional surveys, height may have influenced the decision to seek a karyotype. Consequently, these data cannot serve to implicate increased stature and supernumerary Y chromosomes. However, 31 of the cases include information on height of parents and/or siblings. In 25 of these the 47,XYY proband was taller than his parents and/or siblings. Even so, it remains to be shown that the increased stature of the probands *selected* does not account for their towering above family members.

The Carstairs data showed the 47,XYY men to exceed a control group in height at the .00001 level (Jacobs et al., 1965), and a consecutive admission study at Grendon prison found the two XYYs to be well above a control group of XY admissions (Bartlett et al., 1968). Hook and Kim (1971) reported three 47,XYY boys were taller than a control group ($p < .001$) in an institution for juvenile offenders. Observations from mass screenings indicate normal birth size, but an increase in length becomes apparent shortly after birth.

Information on weight has been provided in 44 case records. Of those 47,XYY males 15 years or older, the weight ranges from 34 kilograms (for a 17-year-old standing 143 centimeters tall) to 111 kilograms (for a 31-year-old standing 203 centimeters). The median weight is 80.1 kilograms.

Although obesity has been noted, and thin body build has also been described, these medians of 186 centimeters tall and 80.1 kilograms represent a medium or normal body build.

### *Ophthalmological Disorders*

Normal eyes have been noted in four cases (Marcus & Richmond, 1970; Parker et al., 1969; Parker et al., 1970; Thorburn et al., 1968). Normal red/green color vision has been cited in four histories (Balodimos et al., 1966; Marcus & Richmond, 1970; Mirouze et al., 1966; Wiener & Sutherland, 1968). Myopia has been cited four times (Court Brown et al., 1968a; Hustinx & van Olphen, 1963; Marinello et al., 1969; Spencer et al., 1969). Cataracts (Lehrnbecher & Lucas, 1969a), corneal opacities (Kosenow & Pfeiffer, 1966), and a discoloration of the lens (Turner & Jennings, 1961, reanalyzed by Court Brown, 1968) have been reported. Iridodenesis (agitated motion of the iris) occurred once (Hustinx & van Olphen, 1963). A mild exophoria (tendency of the eyes to turn outward) in the left eye has been cited (Abdullah et al., 1969).

A specific lack of physical abnormalities has been cited in 15 cases of 47,XYY (Close, Goonetilleke, Jacobs, & Price, 1968; Court Brown et al., 1968a; Cowie & Kahn, 1968; Goodman et al., 1967; Marinello et al., 1969; Sandberg et al., 1961; Spencer et al., 1969; Stenchever & Macintyre, 1969; Wilton & Lever, 1967).

## BEHAVIORAL CHARACTERISTICS

Case reports on 47,XYY males include descriptions ranging from schizophrenic to serious and hard working. Drinking has been noted in eight cases (Balodimos et al., 1966; Daly, 1969b; Griffiths et al., 1970; Marcus & Richmond, 1970; Marinello et al., 1969), but only two cases were addicted. Three cases note drug use without addiction (Griffiths et al., 1970); another case specifically noted its absence (Marcus & Richmond, 1970). Enuresis has been reported in four cases (Abdullah et al., 1969; Leff & Scott, 1968; Persson, 1967; Sandberg et al., 1963).

A wide and heterogeneous collection of descriptive terms is presented in Table 11. Many of these descriptive terms result from clinical impressions. They paint for the most part a picture of seriously disturbed behavior tending toward occasionally severe aggression. However, attempts to assess such aggression with objective indexes have so far failed to produce so striking a result.

At Wandsworth prison, 47,XYY males had significantly more frequent and/or more serious histories of psychiatric treatment than controls ($p < .02$). However, they did not differ on the Eysenck Personality Inventory scales of psychoticism or neuroticism; five of

TABLE 11

BEHAVIORAL CHARACTERISTICS

| Description | Author(s) and year |
|---|---|
| Schizophrenia | Clark, Telfer, Baker, & Rosen, 1970; Daly, 1969b; Falek, Craddick, & Collum, 1970; Marinello, Berkson, Edwards, & Bannerman, 1969 |
| Hallucinations | Abdullah, Jarvik, Kato, Johnston, & Lanzkron, 1969; Daly, 1969a; Marinello et al., 1969 |
| Schizoid affect, flat and labile affect, bland affect | Bartlett, Hurley, Brand, & Poole, 1968; Daly, 1969a; Falek et al., 1970; Marinello et al., 1969 |
| Abnormal shyness, shyness, poor mixer, no close friends | Abdullah et al., 1969; Daly, 1969a; Forssman, Akesson, & Wallin, 1968; Money, Gaskin, & Hull, 1969; Persson, 1967; Wiener & Sutherland, 1968 |
| Inferiority feelings | Leff & Scott, 1968 |
| Suspicious | Daly, 1969a |
| Delusional | Abdullah et al., 1969; Cowling, Rigo, & Martin, 1969; Daly, 1969a |
| Mild depression | Daly, 1969a; Leff & Scott, 1968 |
| Disturbed sleep | Cowling et al., 1969 |
| Childhood or adolescent behavior problem | Abdullah et al., 1969; Falek et al., 1970; Marinello et al., 1969 |
| Sociopath or psychopath | Clark et al., 1970; Daly, 1969b; Marcus & Richmond, 1970; Richards & Stewart, 1966; Wiener, Sutherland, & Bartholomew, 1968 |
| Immature | Abdullah et al., 1969; Daly, 1969a; Falek et al., 1970; Leff & Scott, 1968; Lehrnbecher & Lucas, 1969a; Noel, Quack, Durand, & Rethoré, 1969; Price & Whatmore, 1967a |
| Excessive daydreaming | Money et al., 1969 |
| Impulsive | Borgaonkar, Murdoch, McKusick, Borkowf, Money, & Robinson, 1968; Daly, 1969b; Falek et al., 1970; Marcus & Richmond, 1970; Matthews & Brooks, 1968; Money et al., 1969 |
| Poor judgment | Daly, 1969a; Falek et al., 1970; Marcus & Richmond, 1970 |
| Unrealistic future expectations | Money et al., 1969 |
| Irresponsible, unreliable, sporadic work history | Court Brown, Price, & Jacobs, 1968a; Marcus & Richmond, 1970; Money et al. 1969; Price & Whatmore, 1967b |
| Inadequate | Daly, 1969a; Marinello et al., 1969; Parker, Mavalwala, Weise, Koch, Hatashita, & Cibilich, 1970 |
| Personality disorder | Forssman, Akesson, & Wallin, 1968; Hashi, Tsutsumi, & Tsuda, 1969; Price & Whatmore, 1967a |
| Passive aggressive | Persson, 1967 |
| Low frustration tolerance, easily annoyed | Elbualy, 1969; Persson, 1967 |
| Temper outbursts, tantrums, head-banging tantrums, sudden fits of violent rage and anger, aggressive | Abdullah et al., 1969; Boucharlat & Jalbert, 1969; Clark et al., 1970; Cleveland, Arias, & Smith, 1969; Court Brown et al., 1968a; Cowie & Kahn, 1968; Dallapiccola & Malacarne, 1970; Daly, 1969a; Elbualy, 1969; Forssman, 1967; Forssman et al., 1968; Goodman, Smith, & Migeon, 1967; Heller, 1969; Lisker, Zenzes, & Fonesca, 1968; Marcus & Richmond, 1970; Marinello et al., 1969; Nielsen, Stürup, Tsuboi, & Romano, 1969; Nielsen, Tsuboi, Stürup, & Romano, 1968; Nielsen, Tsuboi, Tuver, Jensen, & Sachs, 1969; Noel, Quack, Durand, & Rethoré, 1969; Parker et al., 1970; Pergament, Sato, Berlow, & Mintzer, 1968; Persson, 1967; Price, Strong, Whatmore, & McClemont, 1966; Rainer, Jarvik, Abdullah, & Kato, 1969; Ricci, 1968; Richards & Stewart, 1966; Thompson, 1968; Turner & Jennings, 1961; Wiener et al., 1968 |
| Restless, tense, anxiety, nervousness | Court Brown et al., 1968a; Cowling et al., 1969; Daly, 1969a; Falek et al., 1970; Leff & Scott, 1968; Marinello et al., 1969; Persson, 1967 |
| Phobias | Cleveland et al., 1969 |
| Obsessional | Forssman et al., 1968; Leff & Scott, 1968 |
| Compulsive | Forssman et al., 1968 |
| Hypochondriacal | Leff & Scott, 1968 |
| Hyperactivity, overactivity, sudden periods of overactivity | Abdullah et al., 1969; Cleveland et al., 1969; Cowie & Kahn, 1968; Daly, 1969a; Parker et al., 1970 |
| Cheerful, superficially cheerful | Marinello et al., 1969; Wiener & Sutherland, 1968 |

Table 11—(*Continued*)

| Description | Author(s) and year |
|---|---|
| Not aggressive, not aggressive to patients, not overtly aggressive | Court Brown et al., 1968a; Goodman et al., 1967; Leff & Scott, 1968 |
| Not a behavior problem, no signs of an antisocial attitude | Daly, Chun, Ewanowski, & Osborne, 1969; Forssman et al., 1968 |
| Passive, rather peaceful, placid and obedient, mild mannered, pleasant and cooperative, gentle manner | Money et al., 1969; Parker, Melnyk, & Fish, 1969; Parker et al., 1970; Rainer et al., 1969; Stenchever & Macintyre, 1969; Thorburn, Chutkan, Richards, & Bell, 1968; Wiener & Sutherland, 1968 |
| Outgoing, friendly, extrovertive with good humor | Cleveland et al., 1969; Daly et al., 1969 |
| Serious, hardworking | Stenchever & Macintyre, 1969 |
| No hallucinations, not delusional | Marcus & Richmond, 1970 |
| Homosexuality | Bartlett et al., 1968; Borgaonkar et al., 1968; Court Brown et al., 1968a; Griffiths, Richards, Zaremba, Abramowicz, & Stewart, 1970; Marcus & Richmond, 1970; Marinello et al., 1969; Melnyk, Thompson, Rucci, Vanasek, & Hayes, 1969; Money et al., 1969; Richards & Stewart, 1966; Wiener & Sutherland, 1968; Wilton & Lever, 1967 |
| No girl friends as adults | Cowling et al., 1969; Griffiths et al., 1970 |
| Never married as adults | Abdullah et al., 1969; Balodimos, Lisco, Irwin, Merrill, & Dignman, 1966; Clark et al., 1970; Goodman et al., 1967; Griffiths et al., 1970; Parker et al., 1969; Price et al., 1966; Wiener & Sutherland, 1968; Wilton & Lever, 1967 |
| Married or common-law wife | Daly, 1969a; Falek et al., 1970; Lisker et al., 1968; Marcus & Richmond, 1970; Sandberg, Koepf, Ishihara, & Hauschka, 1961; Stenchever & Macintyre, 1969; Wilton & Lever, 1967 |
| As a child: Parents separated, broken home, mother died at an early age, deserted by father | Court Brown et al., 1968a; Daly, 1969a; Marinello et al., 1969; Noel, Quack, Durand, & Rethoré, 1969; Ricci & Malacarne, 1964 |

the XYYs tested were all lower on the extraversion scale than their controls (Griffiths et al., 1970).

At Carstairs hospital, 44 psychological variables were assessed, but only 5 showed a significant difference between 47,XYY patients and controls; 2 of those 5 were in the relation of part to whole (Hope, Philip, & Loughran, 1967).

At Grendon prison no clear differences could be shown between two 47,XYY males and a group of controls on the Eysenck Personality Inventory, nor the IPAT anxiety battery, nor a hostility measure (Bartlett et al., 1968).

Clark et al. (1970) reported testing three XYYs with the Minnesota Multiphasic Personality Inventory (MMPI); all three had high scores on Scale 8 (Schizophrenia), and one of the three had a high score on Scale 4 (Psychopathic deviate). Marcus and Richmond (1970) reported the MMPI code type of one XYY to be 4'8. Daly (1969a) cited an MMPI with Scales 8, 4, and 6 (Paranoia) elevated. Falek et al. (1970) presented an 81"472'3960-5 MMPI profile.

The extreme abundance of clinical impressionistic data coupled with the remarkable paucity of psychometric data—and the lack of any definitive findings with what psychometric data have been collected—should serve to call into question the validity of the "aggressive XYY" stereotype. Because of the clear selection bias operating and the high expectancies in many reports, impressionistic data should be especially questioned.

### *Criminal Offenses*

An examination of the offenses listed against these 47,XYY males (Table 12) indicates that among those with criminal records some sort of sexual offense is most common, ranging from indecent exposure when drunk to rape. Theft also is reported frequently, ranging in one case from theft of a milk bottle and a shirt, to car theft in another, to a skillfully planned embezzlement in another.

Price and Whatmore (1967b) examined the records at Carstairs in some detail. They found, however, that the 47,XYY males had *fewer* crimes of violence than a suitably matched control group; 8.7% in the XYY and 21.9% in the controls. In another article (Price & Whatmore, 1967a) they showed that there was very little difference in rate of crimes against property but a considerable difference in crimes against persons: The rate of murder was approximately equal with the XYs only slightly exceeding the XYYs, but the rate of assault was six times higher in XY than XYY males, and the rate of sex assault was two-and-one-half times higher in XY than XYY inmates. These data hardly support the "aggressive XYY" hypothesis. Four of the nine XYY had committed some crime against a person, whereas 17 of the 18 controls had such a record.

Price and Whatmore (1967b) also noted that the XYY had records of convictions earlier than did the controls: Mean age of first conviction was 13.1 years for the XYY and 18 years for the controls ($p < .05$). Daly (1969a) reported a mean age of first conviction of 22.3 for his XYY cases, not significantly different from the controls' 27.2 years. A similar study at Wandsworth prison (Griffiths et al., 1970) failed to support a difference of mean age at first conviction, being 16.7 for the XYY and 16.1 for the XY. Clark et al. (1970) reported a mean age of 12.0 for their XYY cases, more in line with Price and Whatmore than with Daly or Griffiths et al., but these data were partly from juvenile detention centers, and no control data were provided. There are 21 other 47,XYY males with records that cited age at first conviction or detention; they range from 6 to 34 years old with a median of 16, not offering much support for the contention that XYY males generally encounter earlier difficulties with legal authorities (Bartlett et al., 1968; Falek et al., 1970; Hunter, 1968; Leff & Scott, 1968; Nielsen, Tsuboi, Stürup, & Romano, 1968; Noel, Quack, Durand, & Rethoré, 1969; Persson, 1967).

The study at Wandsworth prison (Griffiths et al., 1970) failed to find a difference for mean number of offenses (31 for XYY, 34.5 for XY); nor was there a difference for mean length of total sentences (56 months for XYY, 62.3 months for XY). The mean num-

TABLE 12

CRIMINAL OFFENSES

| Offense | Author(s) and year |
|---|---|
| Sexual | Borgaonkar, Murdoch, McKusick, Borkowf, Money, & Robinson, 1968; Clark, Telfer, Baker, & Rosen, 1970; Court Brown, Price, & Jacobs, 1968a; Daly, 1969a; Griffiths, Richards, Zaremba, Abramowicz, & Stewart, 1970; Marcus & Richmond, 1970; Marinello, Berkson, Edwards, & Bannerman, 1969; Melnyk, Thompson, Rucci, Vanasek, & Hayes, 1969; Nielsen, Tsuboi, Stürup, & Romano, 1968; Persson, 1967; Spencer, Eyles, & Mason, 1969; Telfer, Baker, & Longtin, 1968 |
| Arson or attempted arson | Bartlett, Hurley, Brand, & Poole, 1968; Cowie & Kahn, 1968; Falek, Craddick, & Collum, 1970; Kelly, Almy, & Bernard, 1967; Marinello et al., 1969; Nielsen et al., 1968; Nielsen, Stürup, Tsuboi, & Romano, 1969; Nielsen, Tsuboi, Tuver, Jensen, & Sachs, 1969 |
| Theft | Bartlett et al., 1968; Court Brown et al., 1968a; Daly, 1969a; Forssman, 1967; Griffiths et al., 1970; Leff & Scott, 1968; Marcus & Richmond, 1970; Marinello et al., 1969; Melnyk et al., 1969; Persson, 1967; Wiener, Sutherland, & Bartholomew, 1968 |
| Breaking and entering | Court Brown, 1968; Court Brown et al., 1968a; Daly, 1969a; Griffiths et al., 1970; Marcus & Richmond, 1970 |
| Robbery with violence | Marinello et al., 1969; Persson, 1967 |
| Murder or intent to kill | Clark et al., 1970; Daly, 1969a; Heller, 1969; Wiener et al., 1968 |

ber of times convicted did differ, however, at the .05 level (13 for XYY, 9.2 for controls). The relationship between violence against persons and violence against property is in the same direction as found at Carstairs and again would not seem to support an "aggressive XYY" stereotype.

## INTELLIGENCE

Intelligence level of 47,XYY males has also been a matter of some speculation. Once again the danger of sampling bias is probably disastrous, but of the 75 cases reported, the median IQ is in the 80s with a modal IQ in the 90s (Table 13). The range is from 34 to 125. The distribution of IQ scores is slightly skewed low, but it does not reliably depart from normality ($\chi^2 = 11.64$, $.25 > p > .10$, $df = 9$). At Wandsworth prison, seven of the XYY cases had Wechsler Adult Intelligence Scale (WAIS) scores, and they were matched suitably with controls; the difference favored the controls ($p < .02$; Griffiths et al., 1970). However, Hope et al. (1967) reported that the Carstairs XYY males reflected the distribution of the hospital and did not differ in IQ.

An additional eight cases are cited as average intelligence (Court Brown et al., 1968a; Goodman et al., 1967; Kosenow & Pfeiffer, 1966; Lisker et al., 1968; Marinello et al., 1969; Milcu, Negoescu, Maximilian, Garoiu, Augustin, & Iliescu, et al., 1964; Sandberg et al., 1961; Sandberg et al., 1963). Seven more are reported as being of low intelligence or slow learners (Buckton et al., 1962; Court Brown et al., 1964; Lehrnbecher & Lucas, 1969a; Persson, 1967; Rainer et al., 1969). Three are described as mentally retarded (Castilla, Breg, & Miller, 1964; Fraccaro, Glen Bott, Davies, & Schutt, 1962; Thorburn et al., 1968), three as severely mentally retarded (Fraccaro et al., 1962; Hayward & Bower, 1960; Keutel & Dauner, 1969), and one as mentally defective (Marinello et al., 1969).

These data on intelligence make it seem likely that nearly all ranges of IQ scores are possible in 47,XYY males. Because of selection, however, they do not allow us to conclude that XYY males are likely to be even slightly deficient in intelligence. These data should serve to reinforce curiosity about minor perceptual difficulties that may be found with suitable testing of 47,XYY cases identified from mass screenings.

## CRIMINAL RESPONSIBILITY AND THE 47,XYY

The supposedly aggressive personality of 47,XYY males has been implicated by the defense in court cases in Australia, England, France, the United States, and West Germany (Anonymous—Abnormal Chromosome, 1968; Anonymous—An English XYY Murder Trial, 1969; Anonymous—Born Bad?, 1968; Anonymous—Case of the Double Male, 1969; Anonymous—Extra Chromosome in Court, 1968; Anonymous—Of Chromosomes and Crime, 1968; Anonymous—Question of Y, 1968; Anonymous—XYY Syndrome Ruled No Excuse for Crime, 1969; Bartholomew & Sutherland, 1969; Burns, 1962; Farrell, 1969; Fox, 1969; Graham, 1967; Hall, 1970; Heller, 1969; National Institute of Mental Health, 1970; Wiener, Sutherland, & Bartholomew, 1969). The medico-legal issues

TABLE 13

DISTRIBUTION OF IQ SCORES

| IQ | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | 90–99 | 100–109 | 110–119 | 120–129 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 3 | 2 | 7 | 8 | 11 | 12 | 18 | 9 | 4 | 1 |

Note.—This table is based on reports of Abdullah, Jarvik, Kato, Johnston, & Lanzkron, 1969; Bartlett, Hurley, Brand, & Poole, 1968; Borgaonkar, Murdoch, McKusick, Borkowf, Money, & Robinson, 1968; Boucharlat & Jalbert, 1969; Chaptal, Jean, Bonnet, Emberger, Meylan, & Riev, 1968; Cleveland, Arias, & Smith, 1969; Court Brown, Price, & Jacobs, 1968a; Cowie & Kahn, 1968; Cowling, Rigo, & Martin, 1969; Daly, 1969a; Daly, Chun, Ewanowski, & Osborne, 1969; Dunn, Ford, Auersperg, & Miller, 1961; Elbualy, 1969; Emerit, de Grouchy, Vernant, & Corone, 1967; Falek, Craddick, & Collum, 1970; Forssman, 1967; Forssman, Akesson, & Wallin, 1968; Griffiths, Richards, Zaremba, Abramowicz, & Stewart, 1970; Hashi, Tsutsumi, & Tsuda, 1969; Hunter, 1968a; Leff & Scott, 1968; Lundberg & Wahlstrom, 1970; Marcus & Richmond, 1970; Marinello, Berkson, Edwards, & Bannerman, 1969; Melnyk, Thompson, Rucci, Vanasek, & Hayes, 1969; Papanicolaou, Kirkham, & Loraine, 1968; Parker, Melnyk, & Fish, 1969; Parker, Mavalwala, Weise, Koch, Hatashita, & Cibilich, 1970; Persson, 1967; Price, Strong, Whatmore, & McClemont, 1966; Ricci & Malacarne, 1964; Richards & Stewart, 1966; Shapiro, 1970a; Spencer, Eyles, & Mason, 1969; Turner & Jennings, 1961; Vigi, Volpato, Zardini, Gaburro, & Tiepolo, 1969; Wiener, Sutherland, & Bartholomew, 1968; Wilton & Lever, 1967.

surrounding criminal responsibility and the XYY complement have been discussed at some length (Anonymous—The Defense of Insanity and XYY, 1968; Anonymous—The XYY Chromosome Complement and Criminal Conduct, 1969; Bartholomew, 1968; Baumiller, 1969; Breslau & Bell, 1970; Burke, 1969; Court Brown, 1962, 1967; Elkinton, 1968; Epstein, 1969; Farrell, 1969; Gillette, 1969; Goodhart, 1969; Griffiths & Zaremba, 1967; Kessler & Moos, 1969, 1970; Keutel, 1969; Lehrnbecher & Lucas, 1969b; McWhirter, 1969; Minckler, 1969; Money, 1969; Montague, 1968; Park, 1966; Russell & Bender, 1970; Veylon, 1969). It should be noted that the defense in all of the cases tried involved more by way of psychiatric abnormality than just a 47,XYY karyotype.

## RECOMMENDATIONS

Several recommendations seem to suggest themselves for further study of 47,XYY males. It is often difficult to distinguish new 47,XYY reports from additional information on those previously reported. Some uniform nomenclature must be adopted to identify each XYY example. The Chicago Conference on Standardization in Human Cytogenetics, 1966, recommended an eight-character code consisting of first and last initial, number of day, month, and year of birth. This code has been used by Daly et al. (1969). Whatever identification procedure may be selected, it must be consonant with the ethical principles of confidentiality and privacy (see National Institute of Mental Health Report, 1970).

Cases reported should be more carefully examined for mosaicism. Most investigators report only results of karyotyping with peripheral leukocytes. Once 47,XYY is discovered in blood cells, other tissues should be examined also.

Institutional populations and other select groups, when studied, must be more rigorously specified (Court Brown & Smith, 1969). Population size *must* be given, not just the number selected for study on the basis of some a priori hunch, such as height or kind of criminal record. Where used, selection procedures must be specified in detail.

The bias from noncooperating subjects may be considerable in studies within institutions. Where such bias is likely, it should be candidly reported. Court Brown and Smith (1969) pointed out that in the first Carstairs study (Jacobs et al., 1965), of 203 inmates only 187 agreed to cooperate. Five of these were 47,XYY cases. When 9 of the 16 noncooperators finally agreed, another two 47,XYY cases were found!

Bias may also enter into population studies from failures of tissue cultures from which karyotyping is done. Reports of failure rate run from 1.8% to 2.6% even at best. When only .5% of *all* kinds of chromosome abnormalities are expected, such a bias (if abnormal cells are more likely to fail) could be disastrous.

In the search for sequelae of 47,XYY karyotypes, appropriate sampling techniques must be employed (Suinn, 1969). In the absence of marginal probabilities, one cannot establish the predictability of phenotypic attributes given a 47,XYY constitution from conditional probabilities of 47,XYY given that the phenotypic attribute has been observed. The probability of an Event A, given that an Event B has already been observed is, in general, not equal to the probability of B given A. One need only recall the history pre- and post-1966 of external genital anomaly in 47,XYY to see the crucial nature of this statement. To learn anything definitive about 47,XYY efforts must be made to identify them at birth and institute double blind follow-up studies with suitably matched controls.

The use of the term "YY syndrome" was first suggested by the editors of *Lancet* (Anonymous—The YY Syndrome, 1966). Hienz (1969) has been the most explicit proponent of an XYY syndrome, suggesting that there are four syndrome forms. On the other hand, Boczkowski (1967) argued that in intersexuality such "aetiological syndromes" remain meaningless and misleading, despite the achievements in cytogenetics and biochemistry. He noted that "the question of the existence of the so-called YY syndrome should be reviewed [p. 952]." On the basis of the present review, the use of the term would seem to imply an understanding of phenotypic manifestations of a 47,XYY karyotype that is not in evidence.

The interaction between genotype and environment that produces observable, phenotypic characters is of paramount importance. A given phenotype is not the inevitable consequence of a given genotype. Behavioral geneticists prefer to conceptualize the relative contribution of genetic and environmental factors in determining behavioral outcomes. In the case of 47,XYY males, little can be concluded from the data at hand about phenotypic characters more likely to result from XYY than XY sex chromosomes. Nor are we in a position to assess the 47,XYY constitution's contribution to phenotypic variance for any characteristic.

## REFERENCES

ABDULLAH, S., JARVIK, L. F., KATO, T., JOHNSTON, W. C., & LANZKRON, J. Extra Y chromosome and its psychiatric implications. *Archives of General Psychiatry*, 1969, **21**, 497–501.

AKESSON, H. O., FORSSMAN, H., & WALLIN, L. Chromosomes of tall men in mental hospitals. *Lancet*, 1968, **2**, 1040.

ANONYMOUS. Abnormal chromosome. *British Medical Journal*, 1968, **4**, 398.

ANONYMOUS. An English XYY murder trial. *British Medical Journal*, 1969, **1**, 201.

ANONYMOUS. Anatomy of violence. *Newsweek*, 1970, **75** (January 12), 60–61.

ANONYMOUS. Born bad? *Newsweek*, 1968, **71** (May 6), 87.

ANONYMOUS. Case of the double male. *Newsweek*, 1969, **73** (May 5), 73.

ANONYMOUS. Extra chromosome in court. *Science News*, 1968, **94**, 410.

ANONYMOUS. Of chromosomes and crime. *Time*, 1968, **91** (May 3), 41.

ANONYMOUS. Pursuit of Y. *Newsweek*, 1970, **76** (August 24), 70.

ANONYMOUS. Question of Y. *Time*, 1968, **92** (October 25), 76.

ANONYMOUS. The defense of insanity and XYY. *Australian and New Zealand Journal of Criminology*, 1968, **1**, 199.

ANONYMOUS. The XYY chromosome complement: Brief application to criminal insanity tests. *St. Louis University Law Journal*, 1969, **14**, 297–309.

ANONYMOUS. The XYY chromosome complement and criminal conduct. *Oklahoma Law Review*, 1969, **22**, 287–301.

ANONYMOUS. The YY syndrome. *Lancet*, 1966, **1**, 583–584.

ANONYMOUS. Where have all the XYYs gone? *Lancet*, 1971, **1**, 687.

ANONYMOUS. XYY syndrome ruled no excuse for crime. *Medical World News*, 1969, **10** (15), 23.

BALODIMOS, M. C., LISCO, H., IRWIN, I., MERRILL, W., & DINGMAN, J. XYY karyotype in a case of familial hypogonadism. *Journal of Clinical Endocrinology and Metabolism*, 1966, **26**, 443–452.

BARLOW, P., & VOSA, C. G. The Y chromosome in human spermatozoa. *Nature*, 1970, **226**, 961–962.

BARTHOLOMEW, A. A. The extra Y chromosome and criminal behavior. *Australian and New Zealand Journal of Criminology*, 1968, 2, 6.

BARTHOLOMEW, A. A., & SUTHERLAND, G. A defence of insanity and the extra Y chromosome: R v Hannell. *Australian and New Zealand Journal of Criminology*, 1969, **2**, 29–37.

BARTLETT, D. J., HURLEY, W. P., BRAND, C. R., & POOLE, E. W. Chromosomes of male patients in a security prison. *Nature*, 1968, **219**, 351–354.

BAUMILLER, R. C. XYY chromosome genetics. *Journal of Forensic Sciences*, 1969, **14**, 411–418.

BENDER, M. A., & GOOCH, P. C. An unusually long human Y chromosome. *Lancet*, 1961, **2**, 463–464.

BOCZKOWSKI, K. Describing intersexuality. *Lancet*, 1967, **1**, 952–953.

BOCZKOWSKI, K. Eunuchoidal female with 47,XYY karyotype. *Annales de Genetique*, 1969, **12**, 204–207.

BOCZKOWSKI, K. XYY karyotype in eunuchoidal phenotypic female. *Journal of Clinical Endocrinology and Metabolism*, 1970, **30**, 111–113.

BORGAONKAR, D. S., & HOLLANDER, D. H. Quinacrine fluorescence of the human Y chromosome. *Nature*, 1971, **230**, 52.

BORGAONKAR, D. S., & MULES, E. Comments on patients with sex chromosome aneuploidy: Dermatoglyphics, parental ages, $Xg^a$ blood group. *Journal of Medical Genetics*, 1970, **7**, 345–350.

BORGAONKAR, D. S., MURDOCH, J. L., MCKUSICK, V. A., BORKOWF, S. P., MONEY, J. W., & ROBINSON, B. W. The YY syndrome. *Lancet*, 1968, **2**, 461–462.

BOSAJLEVA, E. Personal communication. Cited by W. M. Court Brown, Males with an XYY sex chromosome complement. *Journal of Medical Genetics*, 1968, **5**, 341–359.

BOUCHARLAT, J., & JALBERT, P. Le triplet chromosomique XYY: Son intérêt en psychiatrie: A propos d'une observation. *Annales Medico-Psychologiques*, 1969, **1**, 793–797.

BRAY, P., & JOSEPHINE, A. An XXXYY sex-chromosome anomaly. *Journal of the American Medical Association*, 1963, **184**, 179–182.

BRESLAU, N. A., & BELL, B. XYY, health and law. *Southern Medical Journal*, 1970, **63**, 831–836.

BUCKTON, K. E., BOND, J. A., & MCBRIDE, J. A. An XYY sex chromosome complement in a male with hypogonadism. Personal communication, 1962. Cited by T. F. J. Hustinx & A. H. F. van Olphen, An XYY chromosome pattern in a boy with Marfan's syndrome. *Genetics*, 1963, **34**, 262–269.

BURKE, K. J. The XYY syndrome: Genetics, behavior, and the law. *Denver Law Journal*, 1969, **46**, 261–284.

BURNS, C. *The Tait case*. Melbourne, Australia: Melbourne University Press, 1962.

CASEY, M. D., BLANK, C. E., STREET, D. R. K., Segall, L. J., McDougall, J. H., McGrath, P. J., & Skinner, J. L. YY chromosomes and antisocial behavior. *Lancet,* 1966, **2**, 859–860.

Casey, M. D., Segall, L. J., Street, D. R. K., & Blank, C. E. Sex chromosome abnormalities in two state hospitals requiring special security. *Nature,* 1966, **209**, 641–642.

Castilla, E. E., Breg, W. R., & Miller, O. J. Personal communication. Cited by O. J. Miller, The sex chromosome anomalies. *American Journal of Obstetrics and Gynecology,* 1964, **90**, 1078-1139.

Cattenach, B. M., & Pollard, C. E. An XYY sex-chromosome constitution in the mouse. *Cytogenetics,* 1969, **8**, 80–86.

Chaptal, J., Jean, R., Bonnet, H., Emberger, J. M., Meylan, F., & Riev, D. Aberration chromosomique de type XYY. *Pédiatrie,* 1968, **53**, 587.

Chu, E. H. Y., & Giles, N. H. Human chromosome complements in normal somatic cells in culture. *American Journal of Human Genetics,* 1959, **11**, 63.

Clark, G. R., Telfer, M. A., Baker, D., & Rosen, M. Sex chromosomes, crime, and psychosis. *American Journal of Psychiatry,* 1970, **126**, 1659–1663.

Cleveland, W. W., Arias, D., & Smith, G. F. Radioulnar synostosis, behavioral disturbance, and XYY chromosomes. *Journal of Pediatrics,* 1969, **74**, 103–106.

Close, H. G., Goonetilleke, A. S. R., Jacobs, P. A., & Price, W. H. The incidence of sex chromosomal abnormalities in mentally subnormal males. *Cytogenetics,* 1968, **7**, 277–285.

Court Brown, W. M. Sex chromosomes and the law. *Lancet,* 1962, **2**, 508–509.

Court Brown, W. M. Genetics and crime. *Journal of the Royal College of Physicians, London,* 1967, **1**, 311–318.

Court Brown, W. M. Males with an XYY sex chromosome complement. *Journal of Medical Genetics,* 1968, **5**, 341–359.

Court Brown, W. M., Buckton, K. E., Jacobs, P. A., Tough, I. M., Kuenssberg, E. V., & Knox, J. D. E. Chromosome studies on adults. *Eugenics Laboratory Monographs,* 1966, No. 42.

Court Brown, W. M., Harnden, D. G., Jacobs, P. A., Maclean, N., & Mantle, D. J. *Abnormalities of the sex chromosome complement in man.* London: Her Majesty's Stationery Office, 1964.

Court Brown, W. M., Price, W. H., & Jacobs, P. A. Further information on the identity of 47,XYY males. *British Medical Journal,* 1968, **2**, 325–328. (a)

Court Brown, W. M., Price, W. H., & Jacobs, P. A. The XYY male. *British Medical Journal,* 1968, **4**, 513. (b)

Court Brown, W. M., & Smith, P. G. Human population cytogenetics. *British Medical Bulletin,* 1969, **25**, 74–80.

Cowie, J., & Kahn, J. XYY constitution in prepubertal child. *British Medical Journal,* 1968, **1**, 748–749.

Cowling, D. C., Rigo, S., & Martin, F. I. R. The "YY syndrome." *Medical Journal of Australia,* 1969, **2**, 443–446.

Dallapiccola, B., & Malacarne, P. Bone abnormalities and XYY syndrome. *Lancet,* 1970, **1**, 311–312.

Daly, R. F. Mental illness and patterns of behavior in 10 XYY males. *Journal of Nervous and Mental Disease,* 1969, **149**, 318–327. (a)

Daly, R. F. Neurological abnormalities in XYY males. *Nature,* 1969, **221**, 472–473. (b)

Daly, R. F., Chun, R. W. M., Ewanowski, S., & Osborne, R. H. The XYY condition in childhood: Clinical observations. *Pediatrics,* 1969, **43**, 852–857.

Davis, R. J., McGee, B. J., Empson, J., & Engel, E. XYY and crime. *Lancet,* 1970, **2**, 1086.

Dent, T., Edwards, J. H., & Delhanty, J. D. A. A partial mongol. *Lancet,* 1963, **2**, 484–487.

Diaso, R. B., & Glass, R. H. The Y chromosome in sperm of an XYY male. *Lancet,* 1970, **2**, 1318–1319.

Dunn, H. G., Ford, D. K., Auersperg, N., & Miller, J. R. Benign congenital hypotonia with chromosomal anomaly. *Pediatrics,* 1961, **28**, 578–591.

Eggen, R. R. Cytogenetics: Review of recent advances in a new field of clinical pathology. *American Journal of Clinical Pathology,* 1963, **39**, 3–37.

Elbualy, M. S. The XYY condition in childhood: Clinical observation. *Pediatrics,* 1969, **44**, 620–621.

Elkinton, J. R. Sex chromosomes and crime. *Annals of Internal Medicine,* 1968, **69**, 399–401.

Emerit, I., de Grouchy, J., Vernant, P., & Corone, P. Chromosomal abnormalities and congenital heart disease. *Circulation,* 1967, **36**, 886–905.

Epstein, C. J. Medical genetics: Recent advances with legal implications. *Hastings Law Journal,* 1969, **21**, 35–49.

Evans, E. P., Ford, C. E., Chaganti, R. S. K., Blank, C. E., & Hunter, H. XY spermatocytes in an XYY male. *Lancet,* 1970, **1**, 719–720.

Evans, E. P., Ford, C. E., & Searle, A. G. A 39, X-41, XYY mosaic mouse. *Cytogenetics,* 1969, **8**, 87–96.

Falek, A., Craddick, R., & Collum, J. An attempt to identify prisoners with an XYY chromosome complement by psychiatric and psychological means. *Journal of Nervous and Mental Disease,* 1970, **150**, 165–170.

Farrell, P. T. XYY syndrome in criminal law: An introduction. *St. John's Law Review,* 1969, **44**, 217–219.

Ferguson-Smith, M. A., Johnston, A. W., & Weinberg, A. N. The chromosome complement in true hermaphroditism. *Lancet,* 1960, **2**, 126–128.

Ford, C. E., & Hamerton, J. L. The chromosomes of man. *Nature,* 1956, **178**, 1020–1023.

Forssman, H. Epilepsy in an XYY man. *Lancet,* 1967, **1**, 1389.

Forssman, H., & Akesson, H. O. Extra Y chromosomes and epilepsy. *Humangenetik,* 1969, **7**, 251–252.

Forssman, H., Akesson, H. O., & Wallin, L. The YY syndrome. *Lancet,* 1968, **2**, 779.

Forssman, H., & Hambert, G. Incidence of Klinefelter's syndrome among mental patients. *Lancet,* 1963, **1**, 1327.

Forssman, H., & Hambert, G. Chromosomes and antisocial behavior. *Lancet,* 1966, **2**, 282.

Forssman, H., & Hambert, G. Chromosomes and antisocial behavior. *Excerpta Criminologica,* 1967, **7**, 113–117.

Fox, R. G. XYY chromosomes and crime. *Australian and New Zealand Journal of Criminology,* 1969, **2**, 5–19.

Fraccaro, M., Glen Bott, M., Davies, P., & Schutt, W. Mental deficiency and undescended testis in two males with XYY chromosomes. *Foila Hereditaria et Pathologica,* 1962, **11**, 211–220. Cited by T. W. J. Hustinx & A. H. F. van Olphen, An XYY chromosome pattern in a boy with Marfan's syndrome. *Genetica,* 1963, **34**, 262–269.

Franks, R. C., Bunting, K. W., & Engel, E. Male pseudohermaphrodism with XYY sex chromosomes. *Journal of Clinical Endocrinology and Metabolism,* 1967, **27**, 1623–1627.

George, K. P. Cytochemical differentiation along human chromosomes. *Nature,* 1970, **226**, 80.

Gillette, P. R. XYY chromosome: Medical and legal aspects. *Science,* 1969, **166**, 947–948.

Goodhart, C. B. XYY chromosome: Legal hazards. *Science,* 1969, **165**, 967.

Goodman, R. M., Miller, J. F., & North, C. Chromosomes of tall men. *Lancet,* 1968, **1**, 1318.

Goodman, R. M., Smith, W. S., & Migeon, C. J. Sex chromosome abnormalities. *Nature,* 1967, **216**, 942–943.

Graham, J. A. M. Chromosomes and crime. *Science Digest,* 1967, **62**, 38–40.

Griffiths, A. W., Richards, B. W., Zaremba, J., Abramowicz, T., & Stewart, A. Psychological and sociological investigation of XYY prisoners. *Nature,* 1970, **227**, 290–292.

Griffiths, A. W., & Zaremba, J. Crime and sex chromosome anomalies. *British Medical Journal,* 1967, **4**, 622.

Hall, G. E. Criminal insanity and the XYY defect. *New Physician,* 1970, **19**, 547–548.

Hashi, N., Tsutsumi, S., & Tsuda, K. The first case report on the XYY syndrome (XYY male) in Japan. *Proceedings of Japanese Academy,* 1969, **45**, 63–67.

Hauschka, T. S., Hasson, J. E., Goldstein, M. N., Koepf, G. F., & Sandberg, A. A. An XYY man with progeny indicating familial tendency to nondisjunction. *American Journal of Human Genetics,* 1962, **14**, 22–30.

Hayward, M. D., & Bower, B. D. Chromosomal trisomy associated with the Sturge-Weber syndrome. *Lancet,* 1960, **2**, 844–846.

Heller, J. H. Human chromosome abnormalities as related to physical and mental dysfunction. *Journal of Heredity,* 1969, **60**, 239–248.

Hienz, H. A. YY-syndrome forms. *Lancet,* 1969, **1**, 155–156.

Hook, E. B., & Kim, D.-S. Prevalence of XYY and XXY karyotypes in 337 nonretarded young offenders. *New England Journal of Medicine,* 1970, **283**, 410–411.

Hook, E. G., & Kim, D.-S. Height and antisocial behavior in XY and XYY boys. *Science,* 1971, **172**, 284–286.

Hope, K., Philip, A. E., & Loughran, J. M. Psychological characteristics associated with XYY sex-chromosome complement in a state mental hospital. *British Journal of Psychiatry,* 1967, **113**, 495–498.

Hsu, L. Y., Shapiro, L. R., & Hirschhorn, K. Meiosis in an XYY male. *Lancet,* 1970, **1**, 1173–1174.

Hudson, B., Burger, H., Wiener, S., Sutherland, G., & Bartholomew, A. A. Plasma testosterone and lutenising hormone in XYY men. *Lancet,* 1969, **2**, 699.

Hultén, M. Meiosis in XYY men. *Lancet,* 1970, **1**, 717–718.

Hultén, M., & Pearson, P. L. Fluorescent evidence for spermatocytes with two Y chromosomes in an XYY male. *Annals of Human Genetics,* 1971, **34**, 273–276.

Hunter, H. Chromatin-positive and XYY boys in approved schools. *Lancet,* 1968, **1**, 816.

Hustinx, T. W. J., & van Olphen, A. H. F. An XYY chromosome pattern in a boy with Marfan's syndrome. *Genetica,* 1963, **34**, 262–269.

Ismail, A. A. A., Harkness, R. A., Kirkham, K. E., Loraine, J. A., Whatmore, P. B., & Brittain, R. P. Effect of abnormal sex-chromosome complements on urinary testosterone levels. *Lancet,* 1968, **1**, 220–222.

Jacobs, P. A., Brunton, M., Melville, M. M., Brittain, R. P., & McClemont, W. F. Aggressive behaviour, mental subnormality, and the XYY male. *Nature,* 1965, **208**, 1351–1352.

Jacobs, P. A., Price, W. H., Court Brown, W. M., Brittain, R. P., & Whatmore, P. B. Chromosome studies on men in a maximum security hospital. *Annals of Human Genetics,* 1968, **31**, 339–347.

Jenkins, R. L. Etiology of mongolism. *American Journal of Diseases of Childhood,* 1933, **45**, 506.

Kelly, S., Almy, R., & Bernard, M. Another XYY phenotype. *Nature,* 1967, **215**, 405.

Kessler, S., & Moos, R. H. XYY chromosome: Premature conclusions. *Science,* 1969, **165**, 442.

Kessler, S., & Moos, R. H. The XYY karyotype and criminality: A review. *Journal of Psychiatric Research,* 1970, **7**, 153–170.

Keutel, J. Forensische Bedeutung des XYY-Status. *Medizinische Klinik,* 1969, **64**, 2251–2257.

Keutel, J., & Dauner, I. XYY-Status bei Kindern. *Zeitschrift für Kinderheilkunde,* 1969, **106**, 314–332.

Kjessler, B. Karyotype, meiosis and spermatogenesis in a sample of men attending an infertility clinic. *Monographs in Human Genetics,* 1966, **2**.

Knox, S. J., & Nevin, N. C. XYY chromosomal constitution in prison populations. *Nature,* 1969, **222**, 596.

Kosenow, W., & Pfeiffer, R. A. YY syndrome with multiple malformations. *Lancet,* 1966, **1**, 1375–1376.

Leff, J. P., & Scott, P. D. XYY and intelligence. *Lancet,* 1968, **1**, 645.

Lehrnbecher, W., & Lucas, G. J. Disorders of brain and connective tissue in a patient with 47,XYY karyotype. *Lancet,* 1969, **2**, 796–797. (a)

Lehrnbecher, W., & Lucas, G. J. XYY chromosome: Medical and legal aspects. *Science,* 1969, **166**, 948–949. (b)

Lisker, R., Zenzes, M. T., & Fonesca, M. T. YY syndrome in a Mexican. *Lancet,* 1968, **2**, 635.

Lubs, H. A., & Ruddle, F. H. Chromosomal abnormalities in the human population: Estimation of rates based on New Haven newborn study. *Science,* 1970, **169**, 495–497.

Lundberg, P. O., & Wahlstrom, J. Hormone levels in men with extra Y chromosomes. *Lancet,* 1970, **2**, 1133.

Maclean, N., Mitchell, J. M., Harnden, D. G., Williams, J., Jacobs, P. A., Buckton, K. A., Baikie, A. G., Court Brown, W. M., McBride, J. A., Strong, J. A., Close, H. G., & Jones, D. C. A survey of sex-chromosome abnormalities among 4514 mental defectives. *Lancet,* 1962, **1**, 293–296.

Marcus, A. M., & Richmond, G. The XYY syndrome: A short review, a case study and investigatory model. *Journal of Forensic Sciences,* 1970, **15**, 154–172.

Marinello, M. J., Berkson, R. A., Edwards, J. A., & Bannerman, R. M. A study of the XYY syndrome in tall men and juvenile delinquents. *Journal of the American Medical Association,* 1969, **208**, 321–325.

Matthews, M. B., & Brooks, P. W. Aggression and the YY syndrome. *Lancet,* 1968, **2**, 355–356.

Mavalwala, J., Parker, C. E., & Melnyk, J. Dermatoglyphics of the XYY syndrome. *American Journal of Physical Anthropology,* 1969, **30**, 209–213.

Mavalwala, J., Thompson, H., Parker, C., & Wilson, G. Sole dermatoglyphics of XYY patients. *American Journal of Physical Anthropology,* 1970, **32**, 145–146.

McIlree, M. E., Price, W. H., Court Brown, W. M., Tulloch, W. S., Newsam, J. E., & Maclean, N. Chromosome studies on testicular cells from 50 subfertile men. *Lancet,* 1966, **2**, 69.

McWhirter, K. XYY chromosome and criminal acts. *Science,* 1969, **164**, 1117.

Melnyk, J., Derencsenyi, A., Vanasek, F., Rucci, A. J., & Thompson, H. XYY survey in an institution for sex offenders and the mentally ill. *Nature,* 1969, **224**, 369–370.

Melnyk, J., Thompson, H., Rucci, A. J., Vanasek, F., & Hayes, S. Failure of transmission of the extra chromosome in subjects with 47,XYY karyotype. *Lancet,* 1969, **2**, 797–798.

Milcu, St. M., Negoescu, I., Maximilian, C., Garoiu, M., Augustin, M., & Iliescu, I. Baiat cu hipospadias si cariotip XYY. *Studii si Ceretări de Endocrinologie,* 1964, **15**, 347. Cited by T. W. J. Hustinx & A. H. F. van Olphen, An XYY chromosome pattern in a boy with Marfan's syndrome. *Genetica,* 1963, **34**, 262–269.

Minckler, L. S. Chromosomes of criminals. *Nature,* 1969, **163**, 1145.

Mirouze, J., Emberger, J.-M., Jaffiol, C., Orsetti, A., & Meynadier, J. Hypopubérisme simple et luxation bilaterale des condes avec aberration chromosomique. (Presence d'un petit acrocentrique supplementaire). *Semaine des Hopitaux de Paris,* 1966, **42**, 753–756.

Money, J. XYY, the law, and forensic moral philosophy. An editorial. *Journal of Nervous and Mental Disease,* 1969, **149**, 309–311.

Money, J., Gaskin, R. J., & Hull, H. Impulse, aggression and sexuality in the XYY syndrome. *St. John's Law Review,* 1969, **44**, 220–235.

Montague, A. Chromosomes and crime. *Psychology Today,* 1968, 2, 42–44; 46–49; 71.

Montero, E., & Duran, J. Polisomia gonosomica XYY. *Annales des Desarrollo,* 1965, **13**, 215–222.

Moor, L. Caryotypes avec deux chromosomes Y et troubles du comportement. *Annales Medico-Psychologiques,* 1967, **1**, 116–126.

National Institute of Mental Health Center for Studies of Crime and Delinquency. *Report on the XYY chromosomal abnormality.* (Public Health Service Publication No. 2103) Chevy Chase, Md.: National Institute of Mental Health, 1970.

Nielsen, J. Y chromosomes in male psychiatric patients above 180 cm. tall. *British Journal of Psychiatry,* 1968, **114**, 1589–1590.

Nielsen, J., Christensen, A., Johansen, S. G., & Frøland, A. Psychopathology and testis histology in a patient with the XYY syndrome. *Acta Medica Scandinavica,* 1966, **180**, 747–757.

Nielsen, J., Johansen, K., & Yde, H. Frequency of diabetes mellitus in patients with Klinefelter's syndrome of different chromosome constitutions and the XYY syndrome. Plasma insulin and growth hormone level after a glucose load. *Journal of Clinical Endocrinology and Metabolism,* 1969, **29**, 1062–1073.

Nielsen, J., & Pedersen, E. Electro-encephalographic findings in patients with Klinefelter's syndrome and the XYY syndrome. *Acta Neurologica Scandinavica,* 1969, **45**, 87–94.

Nielsen, J., Stürup, G., Tsuboi, T., & Romano, D. Prevalence of the XYY syndrome in an institution for psychologically abnormal criminals. *Acta Psychiatrica Scandinavica,* 1969, **45**, 383–401.

Nielsen, J., Tsuboi, T., Stürup, G., & Romano, D. XYY chromosomal constitution in criminal psychopaths. *Lancet,* 1968, **2**, 576.

Nielsen, J., Tsuboi, T., Tuver, B., Jensen, J. T., & Sachs, J. Prevalence and incidence of the XYY syndrome and Klinefelter's syndrome in an institution for criminal psychopaths. *Acta Psychiatrica Scandinavica,* 1969, **45**, 402–424.

Nielsen, J., Yde, H., & Johansen, K. Serum growth hormone level after oral glucose load, urinary excretion of pituitary gonadotropin and 17-ketosteroids in XYY syndrome. *Metabolism,* 1969, **18,** 993–997.

Nielsen, J., Yde, H., Johansen, K. Growth hormone in XYY men. *Lancet,* 1970, **2,** 1963.

Noel, B., Quack, B., Durand, Y., & Rethoré, M.-O. Les hommes 47,XYY. *Annales de Genetique,* 1969, **12,** 223–236.

Noel, B., Quack, B., & Talon, J. Peroxydases leucocytaires et caryotype 47,XYY. *Annales de Genetique,* 1969, **12,** 62–64.

Papanicolaou, A. D., Kirkham, K. E., & Loraine, J. A. Abnormalities in urinary gonadotropin excretion in men with a 47,XYY sex chromosome constitution. *Lancet,* 1968, **2,** 608–610.

Park, W. W. The YY syndrome. *Lancet,* 1966, **2,** 1468.

Parker, C. E. Leutinising hormone in XYY men. *Lancet,* 1969, **1,** 1101.

Parker, C. E., Mavalwala, J., Weise, P., Koch, R., Hatashita, A., & Cibilich, S. The 47,XYY syndrome in a boy with behavior problems and mental retardation. *American Journal of Mental Deficiency,* 1970, **74,** 660–665.

Parker, C. E., Melnyk, J., & Fish, C. H. The XYY syndrome. *American Journal of Medicine,* 1969, **47,** 801–808.

Patau, K. Chromosome identification and the Denver report. *Lancet,* 1961, **1,** 933–935.

Pearson, P. L., & Bobrow, M. Fluorescent staining of the Y chromosome in meiotic stages of the human male. *Journal of Reproduction and Fertility,* 1970, **22,** 177–179.

Pearson, P. L., Bobrow, M., & Vosa, C. G. Technique for identifying Y chromosomes in human interphase nuclei. *Nature,* 1970, **226,** 78–80.

Penrose, L. S. Parental age and mutation. *Lancet,* 1955, **2,** 312.

Penrose, L. S. The causes of Down's syndrome. In D. H. M. Woollam (Ed.), *Advances in teratology.* Vol. 1. London: Logos Press, 1966.

Penrose, L. S. Finger-print pattern and the sex chromosomes. *Lancet,* 1967, **1,** 298–300.

Penrose, L. S., & Loesch, D. A study of dermal ridge width in the second (palmar) interdigital area with special reference to aneuploid states. *Journal of Mental Deficiency Research,* 1967, **11,** 36–42.

Pergament, E., Sato, H., Berlow, S., & Mintzer, R. YY syndrome in an American negro. *Lancet,* 1968, **2,** 281.

Persson, T. An XYY man and his relatives. *Journal of Mental Deficiency Research,* 1967, **11,** 239–245.

Plasse, J. C., Patricot, M. C., Giboulet, M., Noel, B., & Revol, A. L'elimination de la testosterone et des 17-cetosteroides urinaires chez des sujets porteurs d'un caryotype 47,XYY. *Annales de Genetique,* 1970, **13,** 90–95.

Polani, P. E. Chromosome anomalies and abortions. *Developmental Medicine and Child Neurology,* 1966, **8,** 67–70.

Polani, P. E. Abnormal sex chromosomes and mental disorders. *Nature,* 1969, **223,** 680–686.

Price, W. H. The electrocardiogram in males with extra Y chromosome. *Lancet,* 1968, **1,** 1106–1108.

Price, W. H. Sex determination, mental subnormality, crime and delinquency in males. *Journal of Mental Subnormality,* 1969, **15,** 37–44.

Price, W. H., & Jacobs, P. A. The 47,XYY male with special reference to behavior. *Seminars in Psychiatry,* 1970, **2,** 30–39.

Price, W. H., Strong, J. A., Whatmore, P. B., & McClemont, W. Criminal patients with XYY sex-chromosome complement. *Lancet,* 1966, **1,** 565–566.

Price, W. H., & van der Molen, H. J. Plasma testosterone levels in males with the 47,XYY karyotype. *Journal of Endocrinology,* 1970, **47,** 117–122.

Price, W. H., & Whatmore, P. B. Behaviour disorders and pattern of crime among XYY males identified at a maximum security hospital. *British Medical Journal,* 1967, **1,** 533–536. (a)

Price, W. H., & Whatmore, P. B. Criminal behaviour and the XYY male. *Nature,* 1967, **213,** 815. (b)

Rainer, J. D., Jarvik, L. F., Abdullah, S., & Kato, T. XYY karyotype in monozygotic twins. *Lancet,* 1969, **2,** 60.

Ratcliffe, S. G., Stewart, A. L., Melville, M. M., Jacobs, P. A., & Keay, A. Chromosome studies on 3500 newborn male infants. *Lancet,* 1970, **1,** 121.

Recherche Coopérative sur Programme N° 85. Fréquence de la constitution XYY dans la population générale. *Annales de Genetique,* 1968, **11,** 245–246.

Ricci, N. Personal communication. Cited by W. M. Court Brown, Males with an XYY sex chromosome complement. *Journal of Medical Genetics,* 1968, **5,** 341–359.

Ricci, N., & Malacarne, P. An XYY human male. *Lancet,* 1964, **1,** 721.

Richards, B. W., & Stewart, A. The YY syndrome. *Lancet,* 1966, **1,** 984–985.

Rimoin, D. L., & Borgaonkar, D. S. Chromosomal abnormalities in idiopathic osteoarthropathy. *Lancet,* 1966, **2,** 860.

Rudd, B. T., Galal, O. M., & Casey, M. D. Testosterone excretion rates in normal males and males with an XYY complement. *Journal of Medical Genetics,* 1968, **5,** 286–288.

Russell, D. H., & Bender, F. H. Legal implications of the XYY syndrome. *Seminars in Psychiatry,* 1970, **2,** 40–52.

Sandberg, A., Ishihara, T., Crosswhite, L. H., & Koepf, G. F. XYY genotype. *New England Journal of Medicine,* 1963, **268,** 585–589.

Sandberg, A. A., Koepf, G. F., Ishihara, T., & Hauschka, T. S. XYY human male. *Lancet,* 1961, **2,** 488–489.

Santen, R. J., DeKrester, D. M., Paulsen, C. A., & Vorhees, J. Gonadotropins and testosterone in the XYY syndrome. *Lancet,* 1970, **2,** 371.

Sergovich, F., Valentine, G. H., Chen, A. J. C., Kinch, R. A. H., & Smout, M. S. Chromosome aberrations in 2159 consecutive newborn babies.

*New England Journal of Medicine,* 1969, **280,** 851–855.

Shapiro, L. R. Hormones and the XYY syndrome. *Lancet,* 1970, 1, 623. (a)

Shapiro, L. R. Hormones and the XYY male. *Lancet,* 1970, 1, 1347. (b)

Siebner, P., & Pufke, S. Das 47,XYY-syndrom mit berucksichtigung seiner Mosaikformen. *Deutsche Medizinsche Wochenschrift,* 1968, **93,** 2155–2158.

Skakkeback, N. E. Hormones and the XYY male. *Lancet,* 1970, 1, 949–950.

Spencer, D. A., Eyles, J. W., & Mason, M. K. XYY syndrome, and XYY-XXYY mosaicism also showing features of Klinefelter's syndrome. *Journal of Medical Genetics,* 1969, **6,** 159–165.

Stenchever, M. A., & Macintyre, M. N. A normal XYY man. *Lancet,* 1969, **1,** 680.

Suinn, R. M. YY syndrome and sampling techniques. *Lancet,* 1969, **1,** 157–158.

Sumner, A. T., Robinson, J. A., & Evans, H. J. Distinguishing between X, Y and YY-bearing human spermatozoa by fluorescence and DNA content. *Nature New Biology,* 1971, **229,** 231–233.

Sundequist, U., & Hellström, E. Transmission of 47,XYY karyotype? *Lancet,* 1969, **2,** 1367.

Telfer, M., Baker, D., & Longtin, L. YY syndrome in an American negro. *Lancet,* 1968, **1,** 95.

Tettenborn, U., Gropp, A., Murken, J.-D., Tinnefeld, W., Fuhrmann, W., & Schwinger, E. Meiosis and testicular histology in XYY males. *Lancet,* 1970, 2, 267–268.

Tettenborn, U., Schwinger, E., & Gropp, A. [Testicular function in men with XYY chromosomes. Meiosis and morphological findings.] *Deutsche Medizinische Wochenschrift,* 1970, **95,** 158–161.

Thompson, H. Personal communication. Cited by W. M. Court Brown, Males with an XYY sex chromosome complement. *Journal of Medical Genetics,* 1968, **5,** 341–359.

Thompson, H., Melnyk, J., & Hecht, F. Reproduction and meiosis in XYY. *Lancet,* 1967, **2,** 831.

Thorburn, M. J., Chutkan, W., Richards, R., & Bell, R. XYY sex chromosomes in a Jamaican with orthopaedic abnormalities. *Journal of Medical Genetics,* 1968, **5,** 215–219.

Tsuboi, T., & Nielsen, J. Dermatoglyphic study of six patients with the XYY syndrome. *Human Heredity,* 1969, **19,** 299–306.

Turner, B., & Jennings, A. N. Trisomy for chromosome 22. *Archives of Disease in Childhood,* 1961, 2, 49. Cited by W. M. Court Brown, Males with an XYY sex chromosome complement. *Journal of Medical Genetics,* 1968, **5,** 341–359.

Tzoneva-Maneva, T., Bosajleva, E., & Petrov, B. Chromosomal abnormalities in idiopathic osteoarthropathy. *Lancet,* 1966, **1,** 1000–1002.

Uchida, I. A., Ray, M., & Duncan, B. P. 21-trisomy with an XYY sex chromosome complement. *Journal of Pediatrics,* 1966, **69,** 295–298.

Valentine, G. H. The YY chromosome complement. What does it mean? *Clinical Pediatrics,* 1969, **8,** 350–355.

van Wijck, J. A. M., Tijdink, G. A. J., & Stolte, L. A. M. Anomalies in the Y chromosome. *Lancet,* 1962, **1,** 218.

Veylon, R. Chromosome XYY et criminalite. *Presse Medicale,* 1969, **77,** 333–335.

Vigi, V., Volpato, S., Zardini, V., Gaburro, D., & Tiepolo, L. XYY in a ten year old child with aggressive behaviour. *Annales de Genetique,* 1969, **12,** 202–203.

Vignetti, P., Capotori, L., & Ferrante, E. XYY chromosomal constitution with genital abnormality. *Lancet,* 1964, 2, 588–589.

Vosa, C. G. Heterochromatin recognition with fluorochromes. *Chromosoma,* 1970, **30,** 366–372.

Walzer, S., Breau, G., & Gerald, P. S. A chromosome survey of 2400 normal newborn infants. *Journal of Pediatrics,* 1969, **74,** 438–448.

Welch, J. P., Borgaonkar, D. S., & Herr, H. M. Psychopathy, mental deficiency, aggressiveness and the XYY syndrome. *Nature,* 1967, **214,** 500–501.

Wiener, S., & Sutherland, G. A normal XYY man. *Lancet,* 1968, **2,** 1352.

Wiener, S., Sutherland, G., & Bartholomew, A. A. A murderer with 47,XYY and an additional autosomal abnormality. *Australian and New Zealand Journal of Criminology,* 1969, **2,** 20–28.

Wiener, S., Sutherland, G., Bartholomew, A. A., & Hudson, B. XYY males in a Melbourne prison. *Lancet,* 1968, **1,** 150.

Wilton, E., & Lever, A. The YY syndrome. *Lancet,* 1966, **1,** 1156.

Wilton, E., & Lever, A. XYY male. *South African Medical Journal,* 1967, **41,** 284–286.

Zech, L. Investigation of metaphase chromosomes with DNA-binding fluorochromes. *Experimental Cell Research,* 1969, **58,** 463.

(Received January 21, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 3, 234–240

# SOME COMMENTS ON "A BEHAVIORAL APPROACH TO NITROGEN NARCOSIS"

BARRY FOWLER[1]

*Defence Research Establishment Toronto*

R. D. Jennings' recent paper, dealing with nitrogen narcosis, argued that almost no attention has been paid to the behavioral aspects of this phenomenon and that the problem has been approached only from the physiological and medical viewpoint. The present paper reviews a number of behavioral studies of nitrogen narcosis not covered by Jennings and points out certain inaccuracies in his paper. It is concluded that considerable attention has been paid to the behavioral aspects of nitrogen narcosis involving skilled performance but that there is now a need to apply new analytic approaches to this problem.

In a recent review, Jennings (1968) purported to show that a behavioral analysis of nitrogen narcosis has been almost entirely neglected in favor of a physiological and medical approach. He then suggested that a behavioral analysis is necessary and discussed ways such an approach might be profitably applied. Unfortunately, in his review Jennings failed to recognize much of the literature that has been concerned with this kind of analysis. If this literature is taken into account, it becomes evident that the need for a behavioral analysis of nitrogen narcosis has been appreciated and has been pursued side by side with the physiological and medical approach. It is the purpose of this paper to demonstrate this contention by reviewing those studies that were omitted by Jennings and to demonstrate that a number of studies have been concerned with a particular issue raised by Jennings—the relationship between nitrogen narcosis in the field and in the laboratory. A secondary purpose of the present paper is to correct certain inaccuracies in Jennings' paper that are misleading. This secondary purpose is dealt with in the next section, while the two remaining sections are concerned with studies relating to the behavioral analysis of nitrogen narcosis and studies of the relationship between narcosis in the laboratory and in the sea.

## SOME COMMENTS ON SPECIFIC ASPECTS OF JENNINGS' PAPER

Jennings (1968, p. 216) advocated the need for the reduction or elimination of nitrogen narcosis because of the advent of saturation diving. In fact, another inert gas,[2] helium, is invariably used in saturation diving where the pressure is such that nitrogen narcosis would pose a problem. Helium has proven to be nonnarcotic up to the maximum pressures used to date with man. The problem of nitrogen narcosis is more relevant to ordinary diving operations where the use of helium would be uneconomic. For a review of saturation diving, the reader is referred to Chouteau (1969) and MacInnis and Bond (1969).

Jennings (1968, p. 217) referred to a paper by Behnke, Johnson, Poppen, and Motley (1934), and stated that these authors reported the effects of hyperbaric air on performance. In fact, their study was concerned with the effects of hyperbaric *oxygen* and was not directly related to the problem of nitrogen narcosis. Jennings (1968, p. 218) also referred to an abstract by Bennett (1957) (the correct reference is Bennett & Glass,

---

[1] Also at York University. Requests for reprints should be sent to Barry Fowler, Canadian Forces, Personnel Applied Research Unit, Department of National Defence, CFB Toronto, Downsview, Ontario, Canada.

The author gratefully acknowledges the helpful comments of J. Gaito, L. Goldsmith, M. W. Radomski, and L. H. Turl in the preparation of an earlier version of this paper.

[2] The term "inert gas" follows the usage of Featherstone and Muehlbaecher (1963) and refers to any gas that exerts certain *biological* effects without undergoing any change in its chemical structure or modifying the primary chemical structure of other substances. By this definition, anaesthetics such as nitrous oxide and ethylene, as well as the gases helium, neon, nitrogen, argon, krypton, and xenon, are included.

1957), in which it is stated that after alpha blocking had occurred, the answers to an arithmetic problem were both significantly less correct and more delayed. A reading of the full report on which this abstract is based (Bennett & Glass, 1961) reveals that it was not possible to establish a significant relationship between the onset of alpha blocking and performance on the arithmetic task. Jennings (1968 p. 222) implied that the pathologies in the bones (generally called aseptic bone necrosis) of caisson workers may be related to repeated exposures to hyperbaric air. This does appear to be an important factor, but other factors are also involved. The literature on this topic is extensive, going back over many years. Griffiths (1969) and McCallum and Walder (1966) provide recent reviews.

Finally, Jennings reported the pressures used in various experiments in terms of atmospheres, but in some cases failed to convert from the original experimental measure, which was stated in atmospheres absolute (ATA) (e.g., Frankenhaeuser, Graff-Lonnevig, & Hesser, 1963). In the present paper, pressure is reported in terms of ATA (1 ATA = ground level; 2 ATA = 33 feet sea water = 1 atmosphere; 3 ATA = 66 feet sea water = 2 atmospheres).

## ANALYSIS OF BEHAVIOR IN HYPERBARIC AIR

The earliest systematic study was reported by Shilling and Willgrube (1937) who set out to obtain quantitative evidence of the effects of nitrogen narcosis to substantiate the largely descriptive observations that were available at that time. Both highly experienced and trainee divers were used as subjects in a dry compression chamber.[3] They were tested at 10 depths ranging from 1 ATA (control) to 10 ATA on three kinds of tasks: mental arithmetic, digit cancellation, and visual reaction time. Shilling and Willgrube (1937) found a trend toward a decrement in performance as a function of depth on these tasks, particularly on the time to complete the arithmetic and digit cancellation tests, the trainee divers being more affected than the experienced divers. At 10 ATA these decrements were very pronounced, and some subjects were unable to complete the required task. The authors also mentioned emotional changes (particularly euphoria), failure of accomplishment, and memory lapses that occurred in some subjects. They concluded that narcosis results in an increased difficulty in the assimilation of facts and in the making of quick and accurate decisions, together with a lengthening in reaction time and a loss in neuromuscular control and responses.

Case and Haldane (1941) examined mental arithmetic and manual dexterity in a dry compression chamber. Their subjects were not divers, but took part in numerous diving experiments connected with the investigation. Contrary to the Shilling and Willgrube (1937) study, a performance decrement in the arithmetic task only became evident at 8.6 ATA, and the manual dexterity task was found to be affected in only a few subjects at 10 ATA. These authors also emphasized that practical activity and judgment were affected, particularly at 10 ATA. They also observed euphoria and other emotional disturbances at these depths and commented that experience seemed to lead to some adaptation to these effects.

Kiessling and Maag (1962) examined performance on choice reaction time, manual dexterity, and a conceptual reasoning task. They used mostly experienced divers and found impairment on all three tasks at 4 ATA in a dry compression chamber. They disputed the claim by Miles and Mackay (1959) that narcosis does not appear until 7 ATA—a claim made on the basis of finding no performance decrement in experienced divers at 6.5 ATA in the sea on such tasks as mental arithmetic and digit span, forwards and backwards. Kiessling and Maag (1962) also suggested a theoretical rationale for the performance decrements found in hyperbaric environments. This was based on the concept of a hierarchical scheme of behavior organization whereby

> the more recent phylogenetic additions to the nervous system mediate the more complex forms of behavior and are depressed earlier and to a greater

---

[3] A dry compression chamber is filled with pressurized gas, whereas a wet compression chamber is filled with water except for a pressurized layer of gas above the water.

> extent than the older and more primitive forms [p. 92].

They suggested that their results lent some support to this hypothesis, since they found that the percentage decrement in performance on the three tasks was greatest for the conceptual reasoning task and least for the manual dexterity task, and this order was taken by the authors as defining the order of complexity of the tasks on an a priori basis. It should be noted that these authors did not standardize the scores on each test before calculating the percentage decrement in performance, and therefore the comparisons that were made may not be valid.

Adolfson (1967) studied a range of tasks up to 13 ATA using experienced divers and concluded that his results supported the hierarchical organization hypothesis discussed by Kiessling and Maag (1962), since it was found that a mental arithmetic task was more affected than either of two manual dexterity tasks. However, there was no attempt to distinguish between the complexity of the manual dexterity tasks although they were affected differentially and, in addition, the scores on the various tests were not standardized for the purposes of comparison. Adolfson also reported that reaction time increased and the number of responses decreased on a continuous free-association test. These decrements were found to be significantly different from the surface controls at 7 ATA. It was concluded that these findings did not agree with those of Behnke, Thomson, and Motley (1935) whom Adolfson interpreted as finding a decrement in associative reactions at 4 ATA. In fact, Behnke et al. (1935) reported no quantitative results and simply suggested: "A limitation of the power of association and a tendency toward fixation of ideas had to be counteracted [p. 555]." Adolfson also made a number of detailed observations on various aspects of his subjects' behavior at 13 ATA. The subjects were described as showing alterations in mood and affectivity, with euphoria and hypomanic states in some cases. A clouding of awareness and a great variety of perceptual disturbances were reported. In some cases memory was affected. However, it was found that repeated exposures to this depth resulted in disappearance of many of these bizarre symptoms.

Adolfson and Fluur (1967) found that hearing discrimination, as measured by ability to repeat back words heard one at a time through air conduction earphones, decreased by approximately 16% at 11 ATA in a dry compression chamber. They argued:

> The influence of hyperbaric air intoxication on the diver's consciousness and wakefulness and prolonged associative reactions will prevent the sound patterns from being immediately identified when reaching the auditory center and thus will cause a delayed response [p. 175].

In a previous report Fluur and Adolfson (1966), using a standard audiometric technique, had demonstrated a decrease in the auditory threshold at 11 ATA with air conduction earphones, but not with a bone conduction receiver. This finding was interpreted as indicating disturbances of sound conduction through the middle ear. Adolfson and Fluur (1967) considered that the foregoing finding could not explain their results since the words were well above threshold; however, a bone conduction receiver was not used as a control in this experiment.

The question of the minimum pressure at which narcosis can be detected has also been raised by Poulton, Catton, and Carpenter (1964). They used a card-sorting task with caisson workers and found a significant increase in the number of very slow responses at 2 ATA. In addition, at 2.5 ATA variability in timing was increased while at 3.5 ATA a slowing in performance was also evident. When this test was carried out a second time no effect on any indexes of performance was apparent even at 3.5 ATA. This result indicates that hyperbaric air affected performance only while the task was being learned. Poulton et al. (1964) pointed out that these findings raise some doubts about the assumed adequacy of performance by surgeons in pressurized operating theaters.

Some doubt has been cast on these findings, however, as a result of a study by Bennett, Poulton, Carpenter, and Catton (1967). Using the same card-sorting task as Poulton et al. (1964), they found significantly more errors were made at 4 ATA than at the surface breathing air, but they were not able to detect an effect on any index of performance at 2 ATA. Another contradictory finding was that the rate of work was greater at 4 ATA than at the surface, whereas Poulton et al. (1964) found a slowing in the rate of work at 3.5 ATA. Bennett et al. (1967) attributed this increase in the rate of work to an increase in arousal under pressure due to the peculiar conditions that were present in their study. They concluded that 4 ATA is the minimum pressure at which a performance decrement due to nitrogen narcosis is likely to be observed.

### *Discussion*

The foregoing review indicates that Jennings (1968) was incorrect when he concluded that behavioral studies of narcosis have been ignored in the past. His suggestion that more research is required does not focus on important issues that cannot be ignored in any discussion of nitrogen narcosis: These are the degree to which the traditional approach to the analysis of skilled performance while narcotic has proven to be of value and the question of possible fresh approaches to the problem.

The approach adopted in the past has revolved around the use of performance tests, which were seen a priori as representative of the kind of tasks that men under pressure might be called upon to perform in practical situations. Coupled with this has been a concern over the minimum depth at which narcotic effects can be detected. Unfortunately, the selection of tests on the basis of face validity has not proven to be an adequate criterion for the prediction of performance, and at the present time one can only generalize to the extent that manual dexterity tasks are less sensitive than those involving intellectual functions. This difficulty is reflected in the wide range of pressures at which narcosis has been detected. Although some of this variability may be due to differences between groups and motivation, there can be little doubt it also reflects sensitivity differences between ostensibly similar tests.

The approach outlined above has been given theoretical underpinnings by the hierarchical organization hypothesis advanced by Kiessling and Maag (1962). However, this hypothesis has not led to any increase in ability to predict performance. Here, the difficulty lies in the circularity of the definition of the complexity of the task with which the decrement in performance is supposed to correlate. It has not been possible to define complexity other than in terms of relative decrement in performance on the various tasks. The difficulties that this leads to have been discussed above. Further, evidence from studies employing nitrous oxide, where the hierarchical organization hypothesis has been investigated more extensively than in the analysis of nitrogen narcosis, suggests that it would be fruitless to persevere with this approach. Attempts to define complexity on a common-sense a priori basis result in anomalies that cannot be reconciled. Steinberg (1954) found that motor tests were more affected than various associative and relational tasks, although they were hypothesized to be less complex. Similar difficulties have been reported by Frankenhaeuser and Beckman (1961) who found it extremely difficult to differentiate on any rational a priori basis between many tests as to their relative complexity, although the tasks were quite dissimilar.

These criticisms lead to the conclusion that other theoretical approaches to the analysis of the effects of narcosis on performance may be more productive and are deserving of attention. One possibility is the black box model (Welford, 1967), which has been used with some success in the analysis of the effects of aging and of various drugs on performance. Another theoretical approach has been proposed by Reilly and Cameron (1968) and is based on Fleischman's (1967) factor-analytic model of human performance. It remains to be seen whether these or other possible approaches will prove to be more successful than that which has been used traditionally.

## COMPARISON OF NITROGEN NARCOSIS IN THE LABORATORY AND IN THE SEA

Baddeley (1966) tested subjects on a manual dexterity task in a dry compression chamber and in the open sea at 1.3 ATA and 4 ATA. In the pressure chamber, he found a decrement at 4 ATA but not at 1.3 ATA. In the sea, there was a decrement at 1.3 ATA, which was attributed to the various physical difficulties of carrying out the task underwater. However, at 4 ATA in the sea, the size of the decrement was 15% larger than what would be predicted on the basis of the effects of pressure and of working underwater combined. In a further study, Baddeley and Fleming (1967) used the same manual dexterity task as well as a mental arithmetic task but extended the pressure to 7 ATA and used a helium-oxygen mixture as well as air. The same kind of interaction was found for both air and the helium-oxygen mixture, although the size of the effect was less for the latter. In a third experiment (Baddeley, de Figueredo, Hawkswell-Curtis, & Williams, 1968), subjects breathing air at 1.15 and 4 ATA in the sea and in a dry compression chamber were tested on digit copying, a sentence comprehension test, and manual dexterity. In this experiment no interaction was found between pressure and working conditions on any of the measures.

Baddeley (1967) has suggested two possible explanations for the results of these experiments. The first is that anxiety may interact with the partial pressure of nitrogen to increase narcosis. The first two experiments were carried out in the open sea, with conditions on the ocean floor (e.g., temperature and visibility) suboptimal. However, in Experiment 3, the subjects could swim down from a beach to 4 ATA in excellent conditions. Thus, more anxiety is likely to have been generated in the first two experiments than the third. Baddeley argued that the results of Sea Lab II (Bowen, Andersen, & Promisel, 1966) also support his hypothesis since performance in the habitat at 7.2 ATA was unimpaired while performance on tasks outside the habitat was seriously impaired compared to performance in the shallow water control condition. However, these findings do not support Baddeley's hypothesis, since there is no evidence for helium narcosis inside the habitat and, therefore, the difference in performance between the shallow water controls and outside the habitat at 7.2 ATA is most likely due solely to the increased hostility of the environment in the latter condition.

The second possibility suggested by Baddeley (1967) is that anxiety interacts with the stress of working underwater. In support of this contention, Baddeley cited a study by Weltman and Egstrom (1966) who reported that some trainee divers showed a decrement in vigilance performance in shallow water compared to others, although in the control condition on land, it was not possible to differentiate the performance of the two groups. The group that showed the decrement in the underwater condition was identified as more anxious than the other trainees.

There is also a third explanation for Baddeley's results that merits consideration. It is possible that the interaction reported in the first two experiments may be an artifact of the experimental technique. It is unlikely that the shallow water controls replicated exactly the difficult conditions at depth in the first two experiments where, for example, there would be less visibility and it would be colder. Thus, the apparent interaction might only reflect a greater decrement in performance due to the increased hostility of the environment at depth in addition to the decrement due to narcosis. In Baddeley's third experiment (Baddeley et al., 1968), the excellent conditions at all depths meant that this difference between shallow and deep water conditions did not exist to the same extent. Support for this possibility comes from a study by Adolfson (1967) who examined performance on mental arithmetic and manual dexterity to depths of 10 ATA in both a wet and dry compression chamber. In this environment, conditions between the shallow and deep water remain constant. It was found that the decrements attributable to the water and to pressure, respectively, were additive and did not interact. On the other hand, it could be argued that the divers were nonanxious in this situation.

### *Discussion*

The studies reviewed in this section perhaps raise more questions than they answer. Nevertheless, the problem of performance in the undersea environment has not been entirely ignored, as Jennings' (1968) discussion indicated. Most workers interested in nitrogen narcosis are well aware of the need for care in generalizing from the results of pressure chamber studies to open-sea conditions. The major deterrent in initiating studies designed to investigate this problem is the enormous practical difficulties involved in carrying out open-sea experimentation. These difficulties include the need to have available simultaneously suitable support craft, a suitable geographic site, and trained divers.

## REFERENCES

Adolfson, J. *Human performance and behaviour in hyperbaric environments.* Stockholm: Almquist & Wiksell, 1967.

Adolfson, J., & Fluur, E. Hearing discrimination in hyperbaric air. *Aerospace Medicine,* 1967, **38,** 174–175.

Baddeley, A. D. Influence of depth on the manual dexterity of free divers: A comparison between open sea and pressure chamber testing. *Journal of Applied Psychology,* 1966, **50,** 81–85.

Baddeley, A. D. Diver performance and the interaction of stresses. In J. N. Lythgoe & J. D. Woods (Eds.), *Underwater association report 1966–7.* Surrey, England: T.G.W. Industrial and Research Promotions Ltd., 1967.

Baddeley, A. D., de Figueredo, J. W., Hawkswell-Curtis, J. W., & Williams, A. N. Nitrogen narcosis and performance underwater. *Ergonomics,* 1968, **11,** 157–164.

Baddeley, A. D., & Fleming, N. C. The efficiency of divers breathing oxy-helium. *Ergonomics,* 1967, **10,** 311–319.

Behnke, A. R., Johnson, F. S., Poppen, J. R., & Motley, E. P. The effect of oxygen on man at pressures from 1 to 4 atmospheres. *American Journal of Physiology,* 1934, **110,** 565–572.

Behnke, A. R., Thomson, R. M., & Motley, E. P. Psychologic effects from breathing air at four atmospheres pressure. *American Journal of Physiology,* 1935, **112,** 554–558.

Bennett, P. B., & Glass, A. High partial pressures of nitrogen and abolition of blocking of the occipital alpha rhythm. *Journal of Physiology,* 1957, **138,** 18P–19P.

Bennett, P. B., & Glass, A. Electroencephalographic and other changes induced by high partial pressures of nitrogen. *Electroencephalography and Clinical Neurophysiology,* 1961, **13,** 91–98.

Bennett, P. B., Poulton, E. C., Carpenter, A., & Catton, M. J. Efficiency at sorting cards in air and a 20 percent oxygen-helium mixture at depths down to 100 feet and in enriched air. *Ergonomics,* 1967, **10,** 53–62.

Bowen, H. M., Andersen, B., & Promisel, D. Studies of divers' performance during the Sea Lab II project. *Human Factors,* 1966, **8,** 183–199.

Case, E. M., & Haldane, J. B. S. Human physiology under high pressure. *Journal of Hygiene,* 1941, **41,** 225–249.

Chouteau, J. Saturation diving: The conshelf experiments. In P. B. Bennett & D. H. Elliott (Eds.), *The physiology and medicine of diving and compressed air work.* London: Baillière, Tindall & Cassel, 1969.

Featherstone, R. M., & Muehlbaecher, C. A. The current role of inert gases in the search for anaesthesia mechanisms. *Pharmacological Review,* 1963, **15,** 97–121.

Fleischman, E. A. Development of a behavior taxonomy for describing human tasks. *Journal of Applied Psychology,* 1967, **51,** 1–10.

Fluur, E., & Adolfson, J. Hearing in hyperbaric air. *Aerospace Medicine,* 1966, **37,** 783–785.

Frankenhaeuser, M., & Beckman, M. The susceptibility of intellectual functions to a depressant drug. *Scandinavian Journal of Psychology,* 1961, **2,** 93–99.

Frankenhaeuser, M., Graff-Lonnevig, V., & Hesser, C. M. Effects on psychomotor functions of different nitrogen-oxygen gas mixtures at increased ambient pressures. *Acta Physiologica Scandinavica,* 1963, **59,** 400–409.

Griffiths, P. D. Clinical manifestations and treatment of decompression sickness in compressed air workers. In P. B. Bennett & D. H. Elliott (Eds.), *The physiology and medicine of diving and compressed air work.* London: Baillière, Tindall & Cassel, 1969.

Jennings, R. D. A behavioral approach to nitrogen narcosis. *Psychological Bulletin,* 1968, **69,** 216–224.

Kiessling, R. J., & Maag, C. H. Performance impairment as a function of nitrogen narcosis. *Journal of Applied Psychology,* 1962, **46,** 91–95.

MacInnis, J. B., & Bond, G. F. Saturation diving: Man-in-sea and sea lab. In P. B. Bennett & D. H. Elliott (Eds.), *The physiology and medicine of diving and compressed air work.* London: Baillière, Tindall & Cassel, 1969.

McCallum, R. I., & Walder, D. N. Bone lesions in compressed air workers. *Journal of Bone and Joint Surgery,* 1966, **48,** 207–235.

Miles, S., & Mackay, D. E. *The nitrogen narcosis hazard and the self-contained diver.* (Medical Research Council Report. Underwater Physiology Sub-Committee 184 Royal Naval Personnel Research Committee), Alverstoke, Hants, 1959.

Poulton, E. C., Catton, M. J., & Carpenter, A. Efficiency at sorting cards in compressed air. *British Journal of Industrial Medicine,* 1964, **21,** 242–245.

Reilly, R. E., & Cameron, B. J. An integrated measurement system for the study of human performance in the underwater environment. (Office of Naval Research Report No. 196-074) Falls Church, Va.: Biotechnology, Inc., 1968.

Shilling, C. W., & Willgrube, W. W. Quantitative study of mental and neuromuscular reactions as influenced by increased air pressure. *United States Naval Medical Bulletin,* 1937, **35,** 373–380.

STEINBERG, H. Selective effects of an anaesthetic drug on cognitive behaviour. *Quarterly Journal of Experimental Psychology,* 1954, 6, 170–180.

WELFORD, A. T. *Fundamentals of skill.* London: Methuen, 1967.

WELTMAN, G., & EGSTROM, G. H. Perceptual narrowing in novice divers. *Human Factors,* 1966, 8, 499–506.

(Received March 10, 1971)

---

## Manuscripts Accepted for Publication in the *Psychological Bulletin*

The Bogus Pipeline: A New *Ignis Fatuus?* Thomas M. Ostrom (Department of Psychology, 404C West 17th Avenue, Ohio State University, Columbus, Ohio 43210).

Where There is *Ignis,* There May Be Fire. Edward E. Jones (Department of Psychology, Duke University, Durham, North Carolina 27706) and Harold Sigall.

Power Relationships and Group Decision Making in Organizations. Michael T. Wood (Department of Psychology, 404C West 17th Avenue, Ohio State University, Columbus, Ohio 43210).

Socioeconomic Status and Traditional Treatment Approaches Reconsidered. Raymond P. Lorion (Department of Psychology, University of Rochester, Rochester, New York 14627).

Nonnecessary Conditions or Useful Procedures in Desensitization: A Reply to Wilkins. Wesley G. Morgan (Psychological Clinic, Department of Psychology, University of Tennessee, Knoxville, Tennessee 37916).

Gamble's Review of the Holtzman Inkblot Technique: Corrections and Clarifications. Jon D. Swartz (The Hogg Foundation for Mental Health, 310A W. C. Hogg Building, University of Texas, Austin, Texas 78712).

Reactance and the Unwillingness to Help Others. Leonard Berkowitz (Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706).

Moral Conduct and Moral Character: A Psychological Perspective. Robert Hogan (Department of Psychology, The Johns Hopkins University, Baltimore, Maryland 21218).

Direct Visual Perception: A Reply to Gyr. James J. Gibson (Department of Psychology, Social Sciences Building, Cornell University, Ithaca, New York 14850).

Vol. 78, No. 4 October 1972

Psychological Bulletin

Diary No. 734(3)
Date 42
Lib

# PSYCHOSEMANTICS:
# SOME COGNITIVE ASPECTS OF STRUCTURAL MEANING[1]

CHARLES A. PERFETTI[2]

*University of Pittsburgh*

Aspects of semantic structure that have significance for cognitive psychology are discussed. On the one hand, there are developments within the framework of transformational grammar that provide a semantic conception of syntactic relations. On the other hand, there is the structure of lexical meaning, especially the idea of semantic features as components of lexical meaning. While the theoretical status of semantic features is presently weak, there is some evidence that such components function in long-term memory. However, in addition to methodological problems, limitations on semantic features, as they have been previously conceived, are suggested by the possibility of semantic deep structure and the requirements of referential processes.

Much of psycholinguistics has been characterized, until recently, by a focus on syntax. Of course, this is in large part due to the influence of Chomsky and Miller (Chomsky, 1957, 1965; Chomsky & Miller, 1963; Miller, 1962), especially to the initial significance attached to the concept of grammaticality and the concomitant assumption of the separability of semantics from syntax (Chomsky, 1957). Although some psychologists maintained an unwavering interest in meaning, it is a fair assessment to say that a great rush was made to assess transformational grammar as a model for language processing. Against this background of syntactic focus, the present paper discusses semantic structure as a branch of cognitive psychology. The first section deals with the semantics of syntactic structures. The second deals with lexical semantics.

## SYNTACTIC SEMANTICS

The distinction between surface and deep structures implicates meaning. Transformational generative grammar assumes that deep syntactic structure provides the input for semantic interpretation, and, therefore, sentences that have different meaning in general have different deep structures. However, points of departure have emerged within transformational linguistics, centering around the representation of deep structures. It is generally agreed that sentences express certain relationships that must be "deep" in some sense; that is, they are not manifest uniformly in surface structure. Beyond this shared viewpoint, there are by now interesting differences concerning the nature of these underlying relationships.

These different conceptions have come from a number of sources, including McCawley (1968a, 1968b, 1968c, 1970a, 1970b), Fillmore (1968a, 1968b; 1970a, 1970b), Lakoff ([1965],[3] 1968, 1969), and others.[4] Al-

---

[1] Preparation of this paper was supported by the Learning Research and Development Center which is supported in part by funds from the United States Office of Education. The valuable criticisms of James Deese and Alan Lesgold on earlier versions of this paper, which was first written in 1970, are happily acknowledged.

[2] Requests for reprints should be sent to Charles A. Perfetti, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

[3] G. Lakoff. On the nature of syntactic irregularity. Unpublished doctoral dissertation, Indiana University, 1965.

[4] Significant contributions to these developments have come from many sources, including Lyons (1966), Halliday (1966), Bach (1968), Ross [J. R.

© 1972 by the American Psychological Association, Inc.

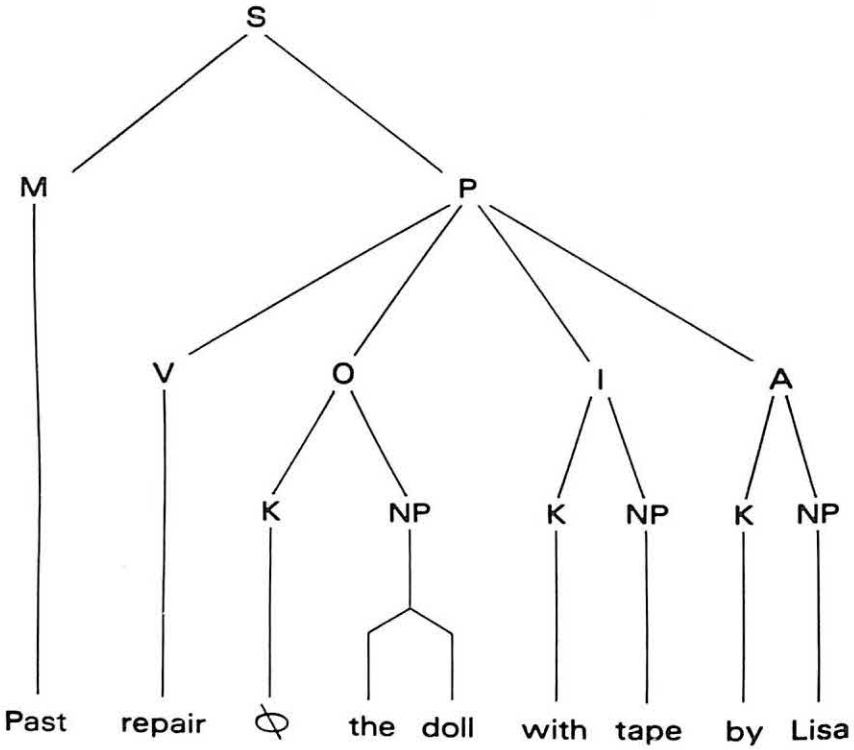


FIG. 1. One possible case grammar representation of the structure underlying the sentence *Lisa repaired the doll with tape*. The cases are Object (O), Instrument (I), and Agent (A). This representation is consistent with Fillmore (1968a) (however, see Footnote 6.)

though in what follows many significant contributions are ignored, the most salient of these developments are represented by brief accounts of the work of Fillmore and McCawley.

*Case Grammar*

Fillmore (1968a) has argued that the base component of language must accommodate the concept of case; that is, base structures contain case relationships as ultimate theoretical terms. Fillmore's position entails several important consequences. For one, deep structures are not generated by phrase structure rewrite rules as they are in standard transformational generative grammar. Rather, unordered case relations are the basic stuff of deep structure, and these relations have syntactic and semantic significance. Furthermore, case relations are only indirectly evident i surface structure. Some languages (e.g Latin) have transformations that yield su face structures that are marked morphologically for case, and others (e.g., English) have transformations that yield surface structures that are more order sensitive and less morphologically marked.

Case relations, although they are treated as syntactic notions, have semantic correspondence. As syntactic descriptions, they replace some fundamental concepts of the standard theory. For example, there is no deep structure division between subject and predicate. A case system recovers the basic relations expressed by the subject-predicate phras structure, but has the additional power o representing important characteristics shared by subject and object.

The heart of a sentence is the propositional component of deep structure. While a second component, *modality*, includes tense, mood

---

Ross. Constraints in variables in syntax. Unpublished doctoral dissertation, M. I. T., 1967], and Postal (1970), among others. The approach represented especially by Lakoff, by Ross, and by McCawley, has been termed "generative semantics."

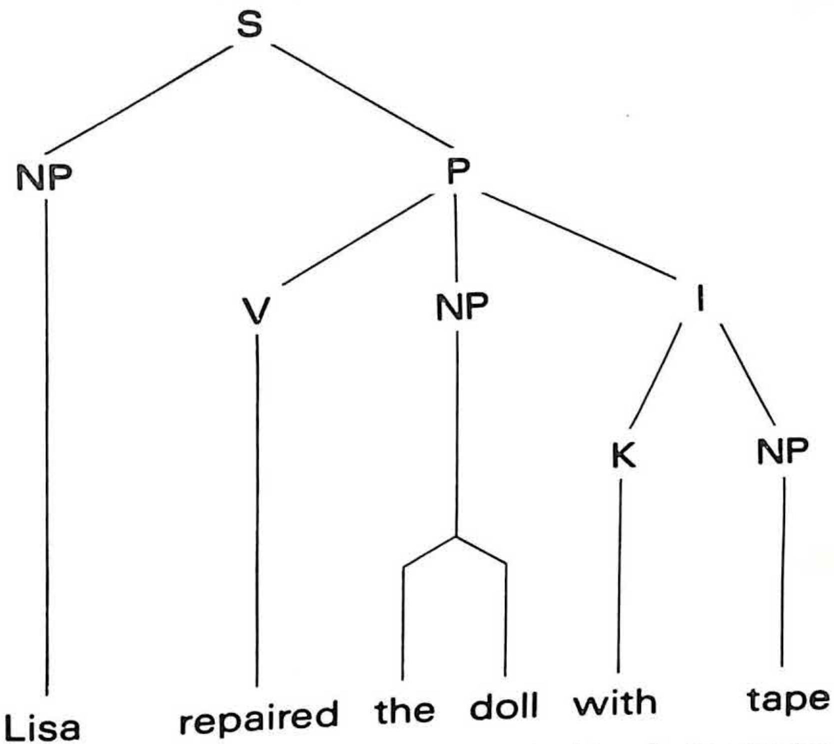


FIG. 2. A possible representation of a sentence derived from the base structure of Figure 1 after a series of transformations.

aspect, etc., the *proposition* is composed of a verb and at least one case category. The list of cases suggested by Fillmore has undergone some change. The most recent list of cases includes Agent, Experiencer, Instrument, Object, Source, Goal, Place, and Time.[5] That these cases are not identical to the ones previously discussed by Fillmore (1968a) indicates that case identification is a difficult empirical problem.

Case grammar deep structure is exemplified in the tree diagram of Figure 1, which employs the notational conventions introduced by Fillmore (1968a). The order of constituents in deep structure is arbitrary, unlike the deep structure of the standard theory. Prepositions are important and are present in deep structure with every noun. From this structure is derived the surface structure shown in Figure 2. Omitted are intermediate structures that would illustrate the series of transformations. These include subject positioning, subject preposition deletion, object preposition deletion, and case label deletion. The transformations are ordered and interdependent, so that, for example, the agent loses its *by* preposition only if it undergoes subject positioning, and case label deletion occurs when a preposition is deleted.[6]

Of course, the agent need not be the surface subject. Thus, in addition to the structure shown in Figure 2, Sentence 1 is derivable from the base of Figure 1.

---

[5] C. J. Fillmore. Some problems for case grammar. Paper presented at Georgetown Roundtable on Linguistics, Georgetown University, March 1971.

[6] This discussion and the structures shown in Figures 1 and 2 are not strictly correct although they are consistent with published treatments. In Fillmore's "Some Problems for Case Grammar" (Paper presented at Georgetown University Roundtable on Linguistics, March 1971), *by* is no longer associated with the agent NP as it is in Figure 1. Rather it is introduced with a passive transformation. More important is the question of how structures are represented. In the Georgetown paper, Fillmore leaves this question unanswered, but discusses the inadequacies of tree diagrams with case labeled nodes of the sort seen in Figure 1. Fillmore has used such representations, however, and they do have the advantage of familiarity, as well as providing the sort of base structure to which transformations are usually applied. The main difficulty is that they represent cases as categories, which they are not, rather than relationships, which they are.

The doll was repaired with tape by Lisa. [1]
The doll was repaired with tape. [2]
The doll was repaired by Lisa. [3]

Sentences 2 and 3 are derived from two different bases, both related to the structure underlying Sentence 1. The main difference is that the structure underlying Sentence 2 has an unspecified agent, while the structure underlying Sentence 3 has an unspecified instrument. The structural similarity can be seen from the facts that *Lisa* is Agent in both Sentence 1 and 3 and *tape* is Instrument in both Sentence 1 and 2.

An important point is that the cases do not have one-to-one correspondence to surface structure relationships. *Lisa* is the same case (Agent) in Sentence 1 as in the sentence of Figure 2, despite surface disparity. (The standard theory represents this state of affairs by assigning *Lisa* as the deep structure noun phrase immediately dominated by S, the sentence node; i.e., *Lisa* is the deep subject of the sentence.) Conversely, in Sentences 2 and 3, *tape* and *Lisa* are involved in similar surface structures, but are different in case category.

While the standard theory's treatment of sentences such as Sentences 2 and 3 can be said to reveal case relations, it can do so only indirectly. For example, if the sentence contains an agent, it is typically the noun phrase immediately dominated by S. Even in Sentence 2, there is such a noun phrase (someone), which, although deleted from the surface structure, as a deep structure subject could be identified with Agent. However, the difference between the standard theory and case grammar goes beyond the explicit marking of case categories. A sentence such as Sentence 4 makes this apparent.

The defendant suffered humiliation at the hands of the prosecutor. [4]

Here, the standard theory shows *defendant* as deep subject, and the Agent role of *prosecutor* is obscured. Thus not only case categories but case relationships are marked in case grammar, and the correspondence between standard deep structure relationships and case relationships is only sometimes revealed. Because the base is unordered, the effect of case relations is to give a degree of abstractness apparently not achieved by the standard theory.

The *propositional* component, which includes case representation, is basically a modified predicate calculus (Reichenbach, 1947). What is conventionally a verb becomes a predicate, and what are conventionally nouns become arguments (Fillmore, 1968b). At least one argument is needed for a predicate (Pa), but in the case of natural languages, two-place predicates are common (Pa,b), and three-place predicates are not rare (Pa,b,c). The following examples show a form of representation based on Fillmore (1968b):

Break (Object). [5]
Hit (Place, Instrument). [6]

The arguments here are generalized case categories that are represented with the verb as inherent arguments. That is, they are arguments that must be stored with the meaning of the verb. They are inherent in a way in which, for example, Instrument is not for *break*. Thus, a particular deep structure involving *break* might be *break* (Object, Instrument) from which could be derived:

The glass was broken by a rock. [7]

as well as

A rock broke the glass. [8]

But because of sentences such as

The glass broke. [9]

*break* is inherently a one-place predicate represented by Example 5. It is important to note that this is a question of deep structure; that is, there must be some sense underlying Sentence 9 in which no Agent or Instrument is involved. Thus for *break*, Instrument and Agent are compatible arguments, but not inherent ones.

By way of comparison, any proposition containing *hit* would require two arguments in deep structure, for example, *hit* (window, rock). This would produce both Sentence 10 and 11.

The window was hit by a rock. [10]
A rock hit the window. [11]

Notice that the requirements that Instrument be an inherent argument prevents

The window hit. [12]

Just as Instrument and Agent are noninherent but compatible arguments for *break*, for *hit*, Agent is compatible. Both of course can enter into three-place arguments:

I $\genfrac{}{}{0pt}{}{\text{hit}}{\text{broke}}$ the window with a rock. [13]

In Fillmore (1968a, 1968b, 1970a, 1970b) there are discussions of various important details of the theory. For one thing, in addition to many of the conventional type transformation rules, rules are needed to select prepositions in surface structure. For example, Instrument must take *with* in the presence of an Agent in surface structure. Also, subject and predicate are strictly surface structure functions in case grammar, and rules for surface function realization of arguments must be associated with each predicate.

The salient feature of case grammar from the present view is the implicit importance of semantics. Fillmore noted that case grammar has been criticized as being "too strongly motivated by semantic considerations." While Fillmore himself insisted on syntactic analysis as the basis of the grammar, he did admit the hope that such analysis reveals important semantic distinctions.

### *Semantic Deep Structure*

A view similar to case grammar with respect to rejecting syntactic deep structure, but different in other respects, comes from McCawley (1968a, 1968c, 1970a, 1970b). The position that syntactic deep structure is unnecessary was earlier suggested by Lakoff and Ross,[7] who share at least some of the views of McCawley.[8]

What the McCawley-Lakoff-Ross positions have in common with Chomsky is (at least) the acceptance of (*a*) a level of semantic representation and (*b*) a level of surface syntactic representation as components of linguistic theory. The point of departure is the question of a level intermediate to Levels *a* and *b* above. McCawley argued that instead of deep structure, a grammar could consist of:

> a "formation-rule component," which specifies the membership of a class of well-formed semantic representations, and a "transformational component" which consists of rules correlating semantic representation [McCawley, 1968c, p. 165].

McCawley makes extensive arguments along a number of different lines. For one, selectional restrictions of the sort proposed by Chomsky (1965) are viewed as derived, not from syntax, but from lexical meaning (McCawley, 1970b). The apparent restrictions on how words may be combined are actually restrictions on how semantic material may be combined—restrictions on "possible messages." The ultimate source of such restrictions derives not from grammar but from the limits of experience and imagination. Thus, McCawley argues that cases in which two words are alleged to have the same meaning but different selection restrictions (which would support a syntactic interpretation) turn out on closer examination to have different meanings. For example, of two different interpretations that could be given to the selectional variation that accompanies *kick* versus *slap*, the position that they both have the same meaning [strike] but with different selection restrictions is rejected. Rather, the semantics of the verbs include presuppositions about type of motion.

To take a syntactic case, selection restrictions prevent the occurrence of sentences such as Sentence 15 as well as Chomsky's (1965) Sentence 14.

The boy frightens sincerity. [14]
He drank the meatloaf. [15]

The standard theory would treat these as violations of syntactic selection restrictions. For example, *frighten* has as a restriction that a noun phrase immediately following it be marked [+ animate]. The restrictions themselves are not at issue. However, McCawley argued such restrictions are imposed by *presuppositions* concerning the denotation of a lexical item (or semantic structure), not by rules of syntax. Use of *frighten* presupposes someone frightenable, and use of *drink* presupposes something drinkable. The effect of this argument is to take away from deep syntax one of its primary functions, namely, to represent selection restrictions, although the argument is not accepted on all sides (cf. Katz, 1970).

---

[7] G. Lakoff & J. R. Ross. Is deep structure necessary? Unpublished manuscript, Cambridge, Mass., 1967.

[8] McCawley gives credit to the unpublished views of George Lakoff for much of what follows especially as presented to the 1967 Texas Conference on Universals. For published views, see Lakoff (1968, 1969). McCawley (1970b) also credits Postal (unpublished; but see Postal, 1970) with having independently arrived at similar conclusions.

A second argument is that a noun phrase must be given an "index," which specifies the reference of a particular noun phrase occurrence. Thus, in the sentence *The marshall chased the outlaw,* some index (e.g., *marshall*$_i$) is attached to *the marshall,* and another index (e.g., *outlaw*$_j$) is attached to *the outlaw*. (The occurrence of indexes as elements of deep structure is indicated more abstractly; e.g., $x_i$ rather than *marshall*$_i$.)

This is not simply an assertion regarding the need for an external (or even internal) referent (cf. Ogden & Richards, 1923). The argument follows Chomsky (1965), who noted that the reflexivization transformation, which converts base structures such as *John hurt John* into *John hurt himself,* must be restricted so as to apply only to noun phrases referring to the same individual. Noun phrase indexes were invented for this purpose, making reflexivization obligatory when indexes are identical and blocked otherwise. McCawley (1970b) has shown that such indexes must be separated from the propositions in which they occur; they are instead constituents of sentences in which the propositions are embedded. The arguments for this claim are based on various cases in which standard phrase structure fails to represent multiple meanings (e.g., McCawley, 1970b, p. 175). For exampe, Sentence 16 has two interpretations.

Ronald has threatened to shoot
a linguist. [16]

This sentence can mean either that Ronald will someday shoot any (indefinite) linguist, or that Ronald is going to shoot a certain person, and the speaker of Sentence 16 describes that person as a linguist. In the latter case *a linguist* is a constituent of the main sentence; in the former case it is a constituent of an embedded sentence underlying *threatened*.

The argument is also made for quantifiers such as *all, every, some,* etc. Quantifiers must have a scope of application indicated in deep structure, as can be seen in the ambiguity of

All boys like all girls. [17]

Sentence 17 can be interpreted as meaning either that every boy has the characteristic of liking all girls or that every girl has the characteristic that all boys like her. In McCawley's proposal, the distribution of quantifiers and indexes would be different for the two cases.

The motivations for these arguments seem to be based on attempts to distinguish meanings and are consistent with at least the spirit of Chomsky (1965). This spirit is one of characterizing underlying meanings. The standard theory does this with a syntactic base component which generates base phrase markers that provide the input for semantic interpretation. McCawley has taken what previously seemed almost a fortuitous by-product of the theory, namely, semantic description, as the basic goal. Hence, where semantic structure and syntactic deep structure appear to diverge, as in the above examples, a deeper level is probed.

A third point is that syntactic categories are realized in surface structure only. Categories of deep structure are similar to those of symbolic logic, although the exact formulation is not identical. (On some reasons for the lack of identity, see McCawley, 1970b, p. 172.) Deep structure consists of a proposition plus at least one index represented as terminal nodes of a semantic tree. For example, in the case of *The marshall chased the outlaw, marshall*$_i$ and *outlaw*$_j$ appear as constituents with *marshall* and *outlaw*, respectively, and with the proposition *chase* as the first constituent. (For arguments that this order is correct for deep structure, i.e., verb first, see McCawley, 1970a.) Although the tree bears a resemblance to syntactic phrase markers, there are substantive differences, especially with respect to what may appear as labels in the tree. Katz (1970) has claimed that such trees are only trivially different from syntactic trees, but this issue would not seem to be merely terminological given the consequences of specific alternative labelings —for example, elimination of labels such as AUX (auxiliary verb), VP (verb phrase), DET (determiner) alters the structure of constituents. Bach (1968) and Lakoff (see Footnote 3) have also argued that syntactic categories are superficial.

The semantic constituent structure goes through intermediate stages. This means, among other things, that the lexical items are the result of transformations on semantic constituents. Thus, for example, the verb *remind* is derived from a deep semantic structure of the form *strike as similar* (Postal, 1970), and some superficially simple verbs such as *kill* are derived from *cause to die* (McCawley, 1968b; in press). Such claims, of course, are not beyond dispute (cf. Fodor, 1970).

McCawley has described his grammar as transformational but not generative, by way of comparison with Chomsky. The same might be said of Fillmore's case grammar. Despite differences, these two approaches share a view of linguistic description that gives a large role to semantics (even if, as in case grammar, supported by "syntactic criteria") in proposing that deep syntax can be done away with.

### *Structure and Process*

While a full evaluation of the significance of controversial developments in linguistics would be premature, these developments may have some significance for psychology. They may provide hypotheses concerning language processing, although there is little research as yet.

One study that has made use of Fillmore's case grammar was carried out by Bobrow (1970) who presented subjects with sentences such as:

The barber pole was on the
covered bridge. [18]

The telephone pole was on the
suspension bridge. [19]

For both sentences, *pole* is Object and *bridge* is Place.

The test is cued recall of the second noun, given the first. In the above example, subjects would have to recall *bridge* given *pole*. The important variation is whether prior to this test, sentence lists presented to the subject contain Sentence 18 once and Sentence 19 once (in a second list) or whether they contain Sentence 18 twice *or* Sentence 19 twice, but not both. If case relations are of major importance for memory, Bobrow reasoned that no difference would be found between the two conditions, since the case relationship between the two nouns was the same, that is, *pole on bridge*. Since only the second noun had to be recalled, the case relation was isolated from other sentence factors. The results were that recall of the second noun was better after two identical sentences than after Sentence 18 and Sentence 19. This supports the importance of identical reference in such a memory process; more particularly, it seems that either a specific or generic reference on a semantic feature complex is encoded.

It must be noted that the importance of case relations was not really examined by Bobrow (aside from the problem of identifying the case hypothesis with the statistical null hypothesis). One ought to change cases without changing references. For example, in addition to Sentence 18 and 19 one could have:

The telephone pole hit the
suspension bridge. [20]

Here reference is the same as Sentence 19 but case is different; *pole* is Instrument and *bridge* is Place (cf. Fillmore, 1968b).

Other implications for cognitive structure come from the representation of lexical information proposed by Fillmore (1968b) and the role of semantic presupposition (Fillmore, 1968b, 1970b). (The latter has entered into Miller's [1969] explanation of sorting data, which is discussed later.) Kintsch (1972) has recently proposed a model of semantic memory that is organized around case grammar principles. Rumelhart, Norman, and Lindsay (1971) have also given a prominent role to case relations in their recent model of semantic memory. By way of support, Kintsch (1972) has found that subjects, without pointed linguistic instructions, will infer implicit information that would formally be represented by suppressed arguments. For example, in *The secretary typed the letter* people offer the observation that the use of typewriter is implied, which supports the notion that Instrument is an inherent argument of *type*. It is a fair assessment, however, to observe that the psycholinguistic implications of case grammar are only beginning to be investigated.

## LEXICAL SEMANTICS

The relational information of the language user is supplemented by information concerning phonological and semantic properties of individual lexical items. For a single word or morpheme, the former are represented by a matrix of distinctive phonetic features and the latter by a matrix of syntactic and semantic features (Chomsky, 1965). It is questions concerning syntactic-semantic features that are one concern of the remainder of this section.[9]

### *Linguistic Features*

Katz and Fodor's (1963) paper on "The Structure of a Semantic Theory" had as its goal the description of the type of semantic theory that would complement the syntactic component of transformational grammar. For this aim, the semantic component required a dictionary to give semantic specification to lexical items and a set of projection rules to yield semantic interpretation of syntactic structures. It is the idea of the dictionary that has had some influence in cognitive psychology.

The dictionary entry for a word was to give systematic account of all its possible senses by means of a branching hierarchy of features. The highest node in a semantic tree was a grammatical marker (e.g., Noun), followed by abstract semantic markers (e.g., Human, Male), in sufficient numbers to distinguish all senses of the word.

The criticisms of this system that have been made by Bollinger (1965), Bierwisch (1969), and Weinreich (1966) bear a brief repeating, especially since semantic features have become psychological concepts. A number of different criticisms point to a central problem: There is no principled limit to the number of semantic markers. This is made especially clear by Bollinger (1965) in his demonstration that the distinction between markers and distinguishers (the terminating node showing the "idiosyncratic" meaning of a word) was totally arbitrary. The crux of the problem is that the number of semantic markers depends on the degree of sensitivity required, and, since the only restriction on the latter is imagination, the semantic structure becomes unstructured.

Another doubtful aspect of this model is the representation of homographs (e.g., *ball*) as single lexical entries, which is a consequence of the theory's emphasis on disambiguation. As Weinreich (1966) noted, disambiguation would seem to be a performance problem, and its large role in a competence model such as Katz and Fodor's is misguided.

Finally, among other problems is the type of information used as semantic markers. While the ability of a speaker to interpret a sentence or to detect an ambiguity depends on nonlinguistic as well as linguistic knowledge, only the latter can be accounted for systematically in a semantic theory. Since the claim regarding the theory's ability to represent linguistic knowledge is questionable, the issue of whether a distinction between linguistic and nonlinguistic knowledge can be supported loses some of its impact.[10]

---

[9] This excludes a good deal of work on meaning, semantic conditioning, and generalization, for example. Creelman (1966) provided a comprehensive survey of this area and of other experimental approaches to meaning (e.g., associative) prior to 1963. Also excluded is the important sort of approach advocated by Rommetveit (1968) who placed meaning in the context of message transmission, and emphasized cognitive, associative, and affective components. Rommetveit urged that we "guard against the Scylla of intuitive speculations and the Charybdis of formally elegant and methodologically sound trivialities [p. 303]." Although with respect to communication, transformational grammar presumably would be up against Charybdis, lexical semantics has been in greater danger from Scylla.

### *Semantic Features in Long-Term Memory*

Although linguistic features of the sort considered above are psychological in that they represent information available to language users, the information is of a very restricted kind. Of broader psychological significance is a conception of semantic features that neither relies exclusively on linguistic criteria nor requires predefined status within one particular linguistic theory. Thus, as a neutral theoretical construct, a semantic feature is an abstract component of the meaning of a lexical item.

The idea that long-term memory includes a representation of semantic features (in the above sense) can be seen in some current models of memory (e.g., Kintsch, 1970; Quillian, 1968, 1969) and in less formal proposals concerning the structure of memory (Bower, 1970; Miller, 1967). Bower (1970) has pointed to general principles of cognitive organization that include the idea that words in memory head a list of features, some of which are semantic. The notion that features represent the information a language user has stored in an internalized dictionary is perhaps the main appeal of features for cognitive psychology.

The empirical support for this notion comes from several sources, most of which are somewhat indirect. The widely observed effect of free recall clustering (Bousfield, 1953; Cofer, 1965; Tulving, 1962) may be interpreted as reflecting organization around shared semantic features. Of course, such organization may be highly subjective, and processes involved are not fully understood (Tulving, 1968). Mandler (1967, 1968) has demonstrated the storage efficiency that results from subjective organization, especially hierarchical organization. Subjects apparently can use hierarchically organized lists, as in taxonomic trees of the type familiar to syntactic and semantic analysis, to increase memorial efficiency (Bower, Clark, Winzenz, & Lesgold, 1969). Furthermore, superordinate cues appear useful to subjects required to retrieve (guess) target words (Zivian & Riegel, 1969).

Such hierarchical organization is an important part of the semantic memory model of Quillian (1968, 1969) and the experiments of Collins and Quillian (1969, 1970). The experiments have studied reaction times for deciding the truth of statements such as (*a*) *a canary is yellow;* (*b*) *a canary has wings;* (*c*) *a canary is a bird;* and (*d*) *a canary is an animal.* It takes less time to decide Statement *a* than Statement *b* and less time to decide Statement *c* than Statement *d* (Collins & Quillian, 1969). The model assumes that properties or features are stored with words in semantic memory. Although the features may be unordered, the words themselves are hierarchically arranged. Thus, *canary* has unordered features which include (sing) and (yellow). The more general feature (wings), however, is not stored with *canary*, but with its superordinate, *bird*. The still more general feature (breathes) would be stored with the higher superordinate, *animal*. It thus takes less time to evaluate Statement *c* than Statement *d*, because the search time must include the additional node from *bird* to *animal*. It takes less time on Statement *a* than on Statement *b* because the relevant information for deciding Statement *a* is stored directly with *bird*, but the information for Statement *b* is obtained after moving up another node. Different models have been proposed by Meyer (1970) and Schaeffer and Wallace (1970) in connection with similar research.

In recognition memory, the role of semantic features has been explicitly investigated, although interpretation is not unequivocal. The fact that false recognition errors tend to occur to words that are semantically related to presented words has been interpreted in terms of semantic features (Anisfeld & Knapp, 1968; Fillenbaum, 1969; Perfetti & Goodman, 1970). Such errors can be seen as resulting from a partial identification of shared semantic features plus a failure to encode a distinguishing feature during either presentation or test. The earlier interpretation of false recognitions as owing to implicit associations (Underwood, 1965) does not seem adequate to explain the fact that these errors occur to synonyms with low associative strength (Anisfeld & Knapp, 1968; Elias & Perfetti[11]; Grossman & Eagle, 1970). Moreover, the fact that false recognitions do not occur to high strength associates of an ambiguous word when the associates are not related to the word's encoded meaning is more readily handled by a semantic feature hypothesis (Perfetti & Goodman, 1970). (Similarly, Hall and Crown [1970] found that errors occurred to associates when they could replace the target word in a sentence, but only infrequently otherwise.)

---

[10] The distinction seems supportable only at a superficial level. For example, judgments of grammaticality can be said to depend on knowledge of the language, but, unless there is something nongeneral about the putative innate component of language, as McNeill (1968) in fact has argued, there is no reason to suppose that such knowledge does not depend on more general cognitive and perceptual capacities in interaction with language and nonlanguage experiences.

Another line of evidence that something like semantic features function in long-term memory comes from word association research. In evaluating a large body of associative data, Deese (1965) concluded that one of the basic associative laws was that words or other cognitive elements are associated when they contrast in a unique and unambiguous way (p. 165). Pollio, Deitchman, and Richards (1969) have proposed a similar principle in terms of shared features.

The principle of contrast proposed by Deese was expressed by McNeill (1966) as the minimal contrast principle. This principle was earlier suggested by Saporta (1959), and has since been explored by several investigators, including Clark (1970), Clifton (1967), Marshall (1968), and Perfetti (1968). According to the minimal contrast principle, two elements will be associated as a function of their shared semantic features, such that association is most probable when the elements form a minimally contrasting pair, that is, contrast on only one feature. A similar proposal comes from Johnson (1970), who argued that associations are determined by distributions of "elementary cognitive characteristics." Semantic features would be one type of elementary cognitive characteristic, which, as a general cognitive element, is free of any arbitrary restrictions on linguistic status.

While several other principles need to be considered, the principle of minimal contrast can account for much association data for words of all grammatical form classes (Clark, 1970). McNeill (1966, 1970), in fact, has proposed that many apparent counter-examples to minimal contrast, for example, syntagmatic associations, can be accounted for by assuming that the contrast occurs on a syntactic category in the absence of sufficiently refined semantic categories. Clark (1970) has pursued a wide range of similar examples, suggesting, among other things, that many associative responses result from matching syntactic selection restrictions with semantic features, for example, *young* → *boy*. This view becomes awkward, of course, if one does away with selection restrictions based on syntax. Such cases then could be thought of as responding with a word whose meaning presupposes the meaning of the stimulus.

The assumption that associations reflect cognitive structures is widely held. The cognitive significance of taxonomic relations, emphasized by Bousfield (1953), has been developed by Riegel (Riegel[12]; Riegel & Riegel, 1963) and extended into several areas, including the use of semantic information in word identification (Zivian & Riegel, 1969), development of semantic skills (Quarterman & Riegel, 1968), semantic structures in bilingualism (Riegel, 1968), and to estimations of meaning similarity measured by overlap of restricted associations (Perfetti, 1967). The related technique of free associative overlap developed by Deese (1962) has been successfully used to display specific semantic relations among animal names (Henley, 1969), as well as to indicate semantic distances among larger sets of words (Deese, 1962, 1965). Whether by free association or restricted association, the semantic structure sought is based on the sort of lexical organization given by semantic features. In addition, the categorical effort provides an empirical suggestion and a limited data base (Riegel, see Footnote 12) for inferring the type of hierarchical relations that are sometimes implied by lexical structure.

---

[11] C. Elias & C. A. Perfetti. Encoding factors in word recognition. Paper presented at the annual meeting of the American Psychological Association, Washington, D. C., September 1971.

[12] K. F. Riegel. *The Michigan restricted association norms.* (Tech. Rep. No. 3) Ann Arbor, Mich.: Department of Psychology, University of Michigan, 1965.

### *Development of Semantic Features*

The development of a semantic system can be seen as involving two aspects of an internalized dictionary. The child acquires new entries for the dictionary, and increases and refines the semantic features for these entries. As McNeill (1970) has observed, this development is typically slow because it depends upon nonsystematic relationships between word forms and underlying meaning structures. The process of adding features to a particular dictionary entry is probably a continuous one, what McNeill (1970) calls "horizontal development," although the added features may be developed elsewhere in the dictionary. McNeill (1966, 1970) has suggested that the shift in children's word associations from syntagmatic to paradigmatic is a reflection of this type of development, that is, the acquisition of semantic features for words that were previously so incompletely structured as to have grammatical class as the only source of contrast. Anderson and Beh (1968) found an increase in paradigmatic recognition errors occurring from age 6 to 7, around the same age as the association shift is observed. Confusing, for example, *light* with *dark* may indicate that some features have been introduced at several places in the dictionary.

This aspect of development is accompanied by some process of generalization in which the acquisition of features is ordered so that later acquired features are increasingly abstract. Anglin (1970) has done a series of experiments that seem to suggest a trend, continuing through adolescence into adulthood, for increasingly abstract features to appear in lexical processes. In word-sorting tasks (Miller, 1969), free recall tasks, and in recall of hierarchically structured material (Bower et al., 1969), there were differences associated with age that indicate that younger children have more idiosyncratic and more concrete lexical organization compared with adults. For example, the minimal feature inferred to be responsible for semantic grouping in a word-sorting task develops along the following lines: (human) → (animal) → (living thing) → (concrete thing) → (entity).

A significant methodological aspect of Anglin's work is that free associative overlap (Deese, 1965) does not show the abstractness trends found with several other methods. Overlap measures may be generally less sensitive to feature differences. Moreover, since semantic structures can be inferred by more direct means (Fillenbaum & Rapoport, 1971; Miller, 1969), there seems little reason for using associative measures for this purpose.

Semantic feature development involves not only the acquisition and refinement of features but also the development of classes and class relations such as *class of* and *same class as*, as well as referential relations, such as *part of*. Riegel (1970) has proposed a model of semantic development based primarily on the growth of such relations. Although it does not use semantic features as theoretical terms, Riegel's theory does postulate abstractions and transformations in semantic processes. Furthermore, the types of class relations it emphasizes would easily accommodate a model of semantic properties of the sort proposed by Quillian (1968) and Collins and Quillian (1969).

### *Theoretical and Methodological Issues*

There are a number of questions associated both with the theoretical status of features and with possible methods of discovery and validation. The consequences of theoretical status include what processes features will explain and what type of discovery procedure is most appropriate. Features such as (± noun) and (+ abstract) are terms of linguistic theory involved in the explanation of certain syntactic processes. Features such as (± pleasant) and (± active) are psychological terms concerning nonsyntactic processes. Following Morris' (1946) trichotomy, one could distinguish three types of features —semantic, syntactic, and pragmatic. The syntactic would include such information as categorial membership (± noun), subcategorial membership (± concrete), and restrictions on the object requirement of a verb

(± _____ noun phrase). The semantic would include denotative features of no more than marginal syntactic significance, for example, (± human) as well as (± liquid). The pragmatic would broadly include nonlinguistic features pertaining to the speaker's relation to the word or its referent, for example, (± anger) and (± pleasant). The addition of pragmatic features is a problem not very different from a distinction between syntactic and semantic features. Both distinctions are easy to see and difficult to support.

Other questions concern the internal structure of a feature representation, especially whether features are ordered. When there is a clear theoretical motivation, the advantage of ordering is significant. A branching hierarchy, for example, allows important generalizations about structure to be represented—for example, that an item which is (+ human) is also (+ animate). In general, a decision for linguistically motivated features seems to entail a decision for order, while the lack of a corresponding order motivation in most psychological theories leaves open questions of order. One alternative to feature hierarchies, previously noted with reference to Quillian (1968, 1969), is to have words, rather than features, hierarchically ordered. This approach has the redundancy advantage of a hierarchy but with unordered features.

Another aspect of internal structure is whether a semantic feature is characterized as a point on a continuum or as a process that can be in a small number of finite states. This may depend in part on theoretical status. For linguistically derived features, the underlying state is discrete, while for pragmatic features, a conception of the underlying state as discrete seems unjustified. *Love* is more (+ pleasant) than *cordiality*, although both are (+ pleasant) with respect to *hostility*.

The discovery and validation of features is still another problem. In principle, features should be discovered, not made. Involved in discovery of features is more than whether a rational or empirical approach is used. The questions concern sophistication versus naiveté, small numbers of informants versus large numbers, and explicit analysis versus implicit analysis. Psychological tradition appears to favor naiveté, large numbers of informants, and implicit analysis. This would suggest a large number of college sophomores rating some measurable linguistic aspect of words or word combinations, and some scaling or factor-analytic method used to display the word groupings and thereby "discover" the features.

As a more general problem, the discovery of features needs to be theoretically motivated. If just those semantic properties that make some difference for syntactic patterns could be discovered, there would be a nonarbitrary limit on a list of features. Otherwise, as Bollinger (1965) demonstrated, there is no limit, since a new feature can always be found that will make a difference for some purpose.

### *Features as Mediators*

Osgood ([1966],[13] [1968][14]) has proposed a treatment of semantic features consistent with his earlier mediational theory of meaning (e.g., Osgood, 1957). The meaning of a word is identified with a "simultaneous bundle of distinctive features." A fragment of the system is illustrated by the following set of feature dimensions: (*a*) moral–immoral; (*b*) potent–impotent; (*c*) active–passive; (*d*) associative–dissociative; (*e*) initiating–reacting; (*f*) ego–alter; and several others. Such dimensions are assumed to be appropriate for interpersonal verbs, for example, *greet, charm, forgive, pardon,* etc. Any verb that has as its denotation an interaction between two or more persons could presumably be coded on these dimensions. Similarly, words naming emotional states can be coded on dimensions of pleasant–unpleasant, active–passive, controlled–uncontrolled, external–internal, and at least eight or nine others (Osgood, see Footnote 13). For example, *disgust* is (unpleasant) and (external), but neither (active) nor (passive) and neither (controlled) nor (uncontrolled), while *horror* is (unpleasant), (active), (uncontrolled), and (external).

---

[13] C. E. Osgood. *Speculation of the structure of interpersonal intentions*. (Tech. Rep. No. 39) Urbana, Ill.: Institute of Communications Research, 1966.

[14] C. E. Osgood. *Interpersonal verbs and interpersonal behavior*. (Tech. Rep. No. 64 (68-9)) Urbana, Ill.: Group Effectiveness Research Laboratory, 1968.

Although the full lexical structure is more involved, the fragment given here can illustrate certain aspects of the approach and problems associated with it:

1. Just as linguistic features belong to a metalanguage, so do psychological features of the sort illustrated here. Thus, there is no necessary verbal connection between words and feature names; for example, between *disgust* and *unpleasant*. Furthermore, while the feature names are attempts to characterize semantic dimensions, the dimension is not transparently related to the names of the features in all cases.
2. Although the fragment shows discrete features, the possibility that a continuum underlies the dimension is left open.
3. The assignment of features to words is subject to empirical validation. One validation procedure is to see how well judgments of acceptability of word combinations yield the same feature groupings as the original codings. In the case of interpersonal verbs, subjects have judged the acceptability of pairs such as *greet humbly, charm ridiculously, forgive ridiculously*. Such procedures apparently have been most successful when applied to judgments of initial sophisticated coders, but not when applied to judgments of college sophomores.

### *Word Sorting*

Miller (1967, 1969) has described a different procedure for discovering semantic features. Subjects are asked to sort words into piles "on the basis of similarity of meaning." There are no restrictions on the number or size of piles that subjects may use. The basic data of this task are the number of subjects who sort any given pair of words into the same pile. The data for all possible word pairs are a symmetrical proximity matrix that can be analyzed in a number of ways, assuming only that the relations among the items can be represented as distances. The choice of subsequent analyses will in part depend on the type of organization holding among the words. Miller (1969) observed that for linearly ordered words (e.g., *baby, child, adolescent, adult*), the sorting method would be inappropriate, and for both linearly ordered words and paradigmatic classes (e.g., kin terms) multidimensional scaling techniques should be used. Otherwise, further analysis is according to Johnson's (1967) program of hierarchical clustering. Both Miller (1969) and Fillenbaum and Rapoport (1971) have used the hierarchical clustering program to determine semantic groupings. The output is a tree graph of words grouped hierarchically beneath nodes determined by either Johnson's (1967) connectedness or diameter method.

Miller's (1969) sample of 48 nouns yielded intuitively plausible structures. For example, apparent clusters of human words emerged (*mother, cook, doctor, umpire, knight*) as did sets of quantitative words (*inch, measure, number, order, scale*). Miller offered a partial explanation for some of the data in terms of *presupposition* and *assertion*. Briefly, an *assertion* is that part of a proposition that normally can be denied, while *presupposition* is that implicit part that is normally not denied. Thus, to say a person is not a mother is normally not to deny that the person is a woman but only that she has children. That is, *mother* presupposes *female*. Miller observed that much, although by no means all, of the clustering data may be explained by supposing that subjects group words together in accordance with their shared presuppositions.

Several procedures for obtaining similarity data, including the sorting technique, have been used by Fillenbaum and Rapoport (1971) for a wide range of words, including color names, kinship terms, pronouns, emotion names, prepositions, conjunctions, verbs of judging, evaluative terms, and verbs of possession. Although these domains represent a much broader semantic range than previously investigated, each domain underwent separate analysis and hence constitutes a homogeneous set in comparison with Miller (1969). The domains vary in the extent of their a priori semantic descriptions. For example, in the case of color names and kinship terms, there are independently hypothesized semantic structures, whereas for evaluative terms and verbs of possession there was little independent semantic analysis. The similarity data for the former type of domain can thus be checked against the independent semantic analyses (e.g., Romney & D'Andrade, 1964, in the case of kinship terms), while for the latter type of domain the outcome is nearer to one of unguided discovery. The qualification on this is that a choice of analytic method imposes a characteristic structure on any set of data. Multidimensional scaling imposes dimensional representation, and hierarchical clustering leads to a taxonomic structure. There is thus a danger of misrepresenting semantic relations, which can be of many types—linear, paradigmatic, nominal, etc., in addition to taxonomic. Nonetheless, there are ways of testing the adequacy of resultant semantic representations, for example, concurrent use of several analytic methods, application of logical preconditions to the choice of method, and validation by intuitive or other independent semantic analyses. Of course, such semantic representations need not correspond to actual cognitive representations, as Deese (1969) has pointed out. They are merely formal representations of semantic relations which people seem to recognize.

The substantive results of this work are considerable. It is noteworthy that in two of the domains, the results favor particular hypothesized semantic structures over others. In the case of kinship terms, the spatial representation (multidimensional scaling) supports the idea that these terms are semantically reciprocal (Romney & D'Andrade, 1964) rather than generated absolutely with ego at the center (Wallace & Atkins, 1960). In the case of personal pronouns, the analysis supports a major semantic distinction between *I* and *other*, rather than among *I*, *you*, *he*. Most of the other analyses reveal interesting semantics, and with the exception of emotional names and evaluative terms, consistent underlying structures with intuitive validity were found.

### *Lexical Markedness*

The idea that parts of the lexicon are comprised of marked and unmarked word pairs has been given prominence by Greenberg (1966), and the cognitive significance of the idea has been developed by Clark (1969a, 1969b)[15] (see also Lyons, 1968, and Marshall, 1968). Following the analogy based on phonology, an unmarked term, but not a marked term, can occur in contexts of neutralization (by definition). That is, in contexts that call for the nominal or nonpolarized sense of an adjective, when only one of a possible pair of adjectives may occur, the one which occurs is the *unmarked*, the nonoccurring adjective, the *marked*. In the sentence *How _____ is the teacher?*, a neutralized context calls for *good, old,* or *interesting,* rather than *bad, young,* or *uninteresting*. If any of the latter group were used, the sense could not be neutral. To ask, "How young?" is to invoke the polarized sense of *young* (i.e., young as opposed to old); but to ask, "How old?" without restrictive context is to invoke the nominal or dimension-naming sense of *old* (i.e., to ask for a value on the dimension of age).

An unmarked adjective thus has a nominal sense and a polarized sense, while a marked adjective has only a polarized sense. The hypothesized cognitive significance of marking hinges on the assumption (Clark, 1969b) that it is the nominal sense of an unmarked adjective which is coded in neutralizable contexts. Thus *old* would be coded as (+ age). Since *young* is coded only as (+ age(−polar)), there is an additional feature to complicate the coding of marked adjectives. Thus, deductive reasoning problems that use marked comparative adjectives require more processing time than parallel problems using unmarked adjectives (Clark, 1969a, 1969b). In Clark's (1969b) examples, *Pete is worse than John* is assumed to entail *John is bad* (hence, polar coding), but *Pete is better than John* is assumed not to entail *John is good* (hence, nominal coding).[16]

---

[15] The development of markedness and neutralization in phonology is owed to Trubetzkoy and *The Prague Circle* (see Vachek, 1964, p. 54f). The generative phonology of Chomsky and Halle (1968) gives a central role to these concepts, as does a treatment of universals by Cairns (1969).

[16] In a study to assess the entailment aspects of marked and unmarked adjectives, subjects were given questions of the form *How wide is x?* and *How narrow is x?* and asked to indicate their willingness to infer whether *x* was wide and narrow, respectively. Subjects made over twice as many inferences from the marked as from the unmarked adjective, in agreement with the markedness hypothesis. However, when comparative sentences similar to those used in Clark's (1969b) experiments were used, the

Similarly in word associations, marked stimuli more often evoke unmarked responses (e.g., *young–old*) than vice versa (*old–young*). Dropping the polarized feature is more likely (cognitively simpler?) than adding it (Clark, 1970; Greenberg, 1966; Marshall, 1968).

The interpretation of such results is open to question. Frequency of usage is closely associated with markedness, such that an unmarked term is always more frequent than the corresponding marked term. Hypotheses concerning differential codability of marked and unmarked terms might be based on frequency as well as semantic complexity, although frequency itself would be in part a function of markedness. Moreover, in the case of word association, the essential assumption that an unmarked adjective be coded in its nominal sense has low plausibility and no independent motivation. Since the word occurs in isolation rather than contexts of neutralization, it may be more plausible to assume that the polarized sense is evoked. More generally, the normal meaning of words such as *good* and *old* may include the polar value. Sentences such as *That's an old house* and *That's a good movie* are at least as normal (unmarked) as questions such as *How old is that house?* or *How good is that movie?* In other words, the polarized use of an adjective may be "unmarked," and its nominal use may be "marked."[17]

## BEYOND FEATURES

Despite the foregoing discussions, there is no reason to suppose that semantic analysis is accomplished simply by a set of semantic features plus interpretive rules for combining the features of individual words. The proposals of McCawley, Fillmore, and others question not only the idea of syntactic deep structure but, necessarily, the interpretive role of semantics. Thus, the very idea of semantic features, which has taken on some force within psychology, is challenged.

Although the full force of the challenge actually includes all the arguments discussed in the first section of this paper, part of the problem can be seen with reference to a putative distinction between feature analysis of lexical items and the structural analysis of "syntactic" constituents. The case can be made that the structural analysis applies as well to words and morphemes (Leech, 1970; McCawley, in press; Weinreich, 1966). For example, words such as *blind* and *laudable* seem as amenable to structural analysis as to feature analysis. Thus, constituents realized as *that cannot see* and *which ought to be praised* can be said to underlie *blind* and *laudable*, respectively (Leech, 1970). More generally, within a lexical item, semantic elements combine in the same way as in larger sentence constituents (Weinreich, 1966). The meaning of *blind* not only contains the meanings of *cannot* and *see*, it contains them in exactly the same relationship as in the paraphrase above. Conversely, the lexical distinctions explained by feature analysis are applicable to constituents as well (Leech, 1970; McCawley, 1970b; Weinreich, 1966). *Buxom lawyer* and *lawyer who is the father of three* show the same contrast as *boy–girl, bull–cow*, etc. Semantic features as abstract components of individual words seems a questionable notion.

One alternative to an interpretive semantics is some sort of *generative* semantics based on an expanded derivative of symbolic logic, as discussed in the first section of this paper. The assumption, common not only to Fillmore, McCawley, and Lakoff but also to Weinreich (1966) and Leech (1970), is that a level of semantic constituent structure exists. (In the case of Weinreich [1966], the constituents are unordered sets of features, called "clusters," that comprise an ordered set of features called a "configuration.")

Lexical features can take on a different look

---

agreement was sharply reduced. Inferences were made in 36% of marked cases and 24% of unmarked cases (Perfetti, unpublished data). It is possible that in comparatives, speakers react to the use of a marked term as an apparent nominal by supposing that, without restrictive context, one should permit a nominal interpretation, despite the fact that it is seldom used nominally—a sort of semantic egalitarianism.

[17] It should be noted that markedness can be considered one among several types of semantic entailment and implication. That is, the relationship of presupposition that holds between *How young is Sam?* and *Sam is young* also holds, for example, between *Sam is effete, even for an intellectual* and *Intellectuals are effete.*

within, such as a generative system. With an interpretive system, features such as (+ animate) are stored with lexical items and are used to restrict the combination of lexical material and to interpret possible meanings. But if, as McCawley (1968c) has argued, selection restrictions are not on how markers and features are combined, then this type of feature is unnecessary. Of course, semantic properties of some sort are clearly necessary for interpretation, and one does not abandon the position that morphemes are not the ultimate units of meaning.

Fillmore (1968b, 1970b) takes a similar position. The lexicon contains a basic sense characterization for each word, but this has no particular theoretical status. What does have theoretical importance is the remainder of the lexical entry for each word: (*a*) the inherent arguments, (*b*) the interpretation restriction on inherent arguments, (*c*) implicit arguments, (*d*) compatible noninherent arguments, (*e*) case suppression, (*f*) surface structure relations of arguments, (*g*) surface structure preposition selection, and a few other facts. Some of these have been dealt with in the first section of this paper. Here, the point is that what is stored in the lexicon is not semantic features, but entries governing the use of a word based on its assumed *denotation*. To illustrate just one aspect, Fillmore, like McCawley, assumes that anomaly arises when semantic presuppositions are not met, and these are represented with the lexical entry, *b* above. For example, a normal interpretation of *break* restricts it to objects that are rigid. There is no linguistic feature such as (+ rigid) involved. Rather there is a restriction on the inherent arguments of *break*, namely, Object, that they *denote* things that are characteristically rigid. The strangeness of the sentence *She broke her handkerchief* is therefore an interpretation problem: The normal interpretation of either *broke* or *handkerchief* is altered.

One of the factors that emerges as significant is the role of reference. The argument of McCawley (1970b) on this point has already been noted, and in Fillmore (1968a, 1968b, 1970b) there is much importance implicitly placed on reference. Of course, reference is a difficult problem, and linguistic theory, even generative semantics, is not its natural habitat. Olson (1970) has recently discussed theories of reference, including the well-known meaning triangle of Ogden and Richards (1923). Olson reviewed the shortcomings of this and other ideas of reference and offered a theory that emphasizes that reference is a differentiation among a set of alternatives—essentially an explanation of psychological context of the sort not considered by linguistics. Meaning then is conceived as the information that differentiates among the perceived or inferred set of alternatives, and Olson (1970) reported data that confirm that actual utterances which communicate information depend on what the alternatives are, even though the same physical fact is being communicated in each case. There is a little doubt that a speaker's intention with respect to a presupposed range of alternatives partly determines choice of expression. However, neither is there any doubt that semantic processing is not in principle dependent upon such factors. Sentences are understood in isolation. Nonetheless, like McCawley, Olson argued explicitly against syntactic deep structure, and, by implication, against semantic features.

Recent experiments by Anisfeld (1970) seem to suggest that when possible, reference rather than semantic features is encoded. The false recognitions that may be based on semantic feature overlap were found for contrasting adjectives, for example, *thick* and *thin*. But when noun phrases that contain such adjectives were tested, recognition confusions occurred only for noun phrases of similar denotation. For example, confusions between *thick book* and *thin book* did not occur. What did occur were confusions between *thick book* and *fat book*. Anisfeld (1970) argued that adjectives alone are less likely to be referentially encoded, and contrasts are confusable on the basis of shared features. However, the indefinite reference of an isolated adjective becomes part of more definite reference when it is part of a noun phrase. In view of such results, it would be useful to look at other form-class differences in recognition errors, as well as other factors associated with reference, such as concreteness. The study of reference thus brings us closer not only to the mental operations involved in imagery (Begg & Paivio, 1969; Paivio, 1969) but also closer to a cognitive semantics and farther from the pristine linguistic features involved in the semantic component of transformational grammar.

## REFERENCES

Anderson, S. A., & Beh, W. The reorganization of verbal memory in childhood. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 1049–1053.

Anglin, J. M. *The growth of word meaning.* Cambridge, Mass.: M. I. T. Press, 1970.

Anisfeld, M. False recognition of adjective-noun phrases. *Journal of Experimental Psychology,* 1970, **86**, 120–122.

Anisfeld, M., & Knapp, M. Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology,* 1968, **77**, 171–179.

Bach, E. Nouns and noun phrases. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory.* New York: Holt, Rinehart & Winston, 1968.

Begg, I., & Paivio, A. Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 821–827.

Bierwisch, M. On certain problems of semantic representation. *Foundations of Language,* 1969, **5**, 153–217.

Bobrow, S. A. Memory for words in sentences. *Journal of Verbal Learning and Verbal Behavior,* 1970, **9**, 363–372.

Bollinger, D. The atomization of meaning. *Language,* 1965, **41**, 553–573.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology,* 1953, **49**, 229–240.

Bower, G. H. Organizational factors in memory. *Cognitive Psychology,* 1970, **1**, 18–46.

Bower, G. H., Clark, M., Winzenz, D., & Lesgold, A. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 323–343.

Cairns, C. E. Markedness, neutralization, and universal redundancy rules. *Language,* 1969, **45**, 863–885.

Chomsky, N. *Syntactic structures.* The Hague: Mouton, 1957.

Chomsky, N. *Aspects of the theory of syntax.* Cambridge, Mass.: M. I. T. Press, 1965.

Chomsky, N., & Halle, M. *The sound pattern of English.* New York: Harper & Row, 1968.

Chomsky, N., & Miller, G. A. Introduction to the formal analysis of natural languages. In R. D. Luce, R. B. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology.* Vol. 2. New York: Wiley, 1963.

Clark, H. H. Influence of language in solving three-term series problems. *Journal of Experimental Psychology,* 1969, **82**, 205–215. (a)

Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review,* 1969, **76**, 387–404. (b)

Clark, H. H. Word associations and linguistic theory. In J. Lyons (Ed.), *New horizons in linguistics.* Baltimore: Pelican Books, 1970.

Clifton, C. The implications of grammar for word associations. In K. Salzinger & S. Salzinger (Eds.), *Research in verbal behavior and some neurophysiological implications.* New York: Academic Press, 1967.

Cofer, C. N. On some factors in the organizational characteristics of free recall. *American Psychologist,* 1965, **20**, 261–272.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 240–247.

Collins, A. M., & Quillian, M. R. Does category size affect categorization time? *Journal of Verbal Learning and Verbal Behavior,* 1970, **9**, 432–438.

Creelman, M. M. *The experimental investigation of meaning.* New York: Springer, 1966.

Deese, J. On the structure of associative meaning. *Psychological Review,* 1962, **69**, 161–175.

Deese, J. *The structure of associations in language and thought.* Baltimore: Johns Hopkins Press, 1965.

Deese, J. Conceptual categories in the study of content. In G. Gerbner (Ed.), *The analysis of communication content.* New York: Wiley, 1969.

Fillenbaum, S. Words as feature complexes: False recognition of antonyms and synonyms. *Journal of Experimental Psychology,* 1969, **82**, 400–402.

Fillenbaum, S., & Rapoport, A. *Structure in the subjective lexicon.* New York: Academic Press, 1971.

Fillmore, C. J. The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory.* New York: Holt, Rinehart & Winston, 1968. (a)

Fillmore, C. J. Lexical entries for verbs. *Foundations of Language,* 1968, **4**, 373–393. (b)

Fillmore, C. J. The grammar of *hitting* and *breaking.* In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English transformational grammar.* Waltham, Mass.: Ginn, 1970. (a)

Fillmore, C. J. Types of lexical information. In F. Kiefer (Ed.), *Studies in syntax and semantics.* Dordrecht, Holland: D. Reidel, 1970. (b)

Fodor, J. A. Three reasons for not deriving "kill" from "cause to die." *Linguistic Inquiry,* 1970, **1**, 429–438.

Greenberg, J. H. *Language universals.* The Hague: Mouton, 1966.

Grossman, L., & Eagle, M. Synonymity, antonymity, and association in false recognition responses. *Journal of Experimental Psychology,* 1970, **83**, 244–248.

Hall, J. W., & Crown, I. Associative encoding of words in sentences. *Journal of Verbal Learning and Verbal Behavior,* 1970, **9**, 303–307.

Halliday, M. A. K. Some notes on "deep" grammar. *Journal of Linguistics,* 1966, **2**, 55–67.

Henley, N. M. A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 176–184.

Johnson, M. G. A cognitive-feature model of compound free associations. *Psychological Review,* 1970, **77**, 282–293.

Johnson, S. C. Hierarchical clustering schemes. *Psychometrika,* 1967, **32**, 241–254.

Katz, J. J. Interpretive semantics vs. generative semantics. *Foundations of Language,* 1970, **6**, 220–259.

Katz, J. J., & Fodor, J. A. The structure of a semantic theory. *Language,* 1963, **39**, 170–210.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory,* New York: Academic Press, 1972.

Lakoff, G. Instrumental adverbs and the concept of "deep structure." *Foundations of Language,* 1968, **4**, 4–29.

Lakoff, G. Presupposition and relative grammaticality. *Studies in Philosophical Linguistics,* 1969, **1**, 103–116.

Leech, G. N. *Towards a semantic description of English.* Bloomington, Ind.: Indiana University Press, 1970.

Lyons, J. Towards a notional theory of the parts of speech. *Journal of Linguistics,* 1966, **2**, 209–236.

Lyons, J. *Introduction to theoretical linguistics.* Cambridge, England: Cambridge University Press, 1968.

Mandler, G. Organization and memory. In K. W. Spence & J. A. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

Mandler, G. Association and organization: Facts, fancies, and theories. In T. R. Dixon & D. C. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Marshall, J. Psychological aspects of semantic structure. In R. Meetham (Ed.), *Encyclopedia of linguistics, information and control.* London: Pergamon Press, 1968.

McCawley, J. D. Concerning the base component of the transformational grammar. *Foundations of Language,* 1968, **4**, 243–269. (a)

McCawley, J. D. Lexical insertion in a transformational grammar without deep structure. *Papers from the 1968 Chicago Linguistic Society Regional Meeting,* 1968, 71–80. (b)

McCawley, J. D. The role of semantics in grammar. In E. Bach & R. R. Harms (Eds.), *Universals in linguistic theory.* New York: Holt, Rinehart & Winston, 1968. (c)

McCawley, J. D. English as a VSO language. *Language,* 1970, **46**, 286–299. (a)

McCawley, J. D. Where do noun phrases come from? (Unpublished revised version of paper). In R. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English transformation grammar.* Waltham, Mass.: Ginn, 1970. (b)

McCawley, J. D. Semantic representation. In P. Garvin (Ed.), *Cognition: A multiple view.* New York: Spartan Books, in press.

McNeill, D. A. A study of word association. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 548–557.

McNeill, D. On theories of language acquisition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

McNeill, D. *The acquisition of language.* New York: Harper and Row, 1970.

Meyer, D. E. On the representation and retrieval of stored semantic information. *Cognitive Psychology,* 1970, **1**, 242–300.

Miller, G .A. Some psychological studies in grammar. *American Psychologist,* 1962, **17**, 748–762.

Miller, G. A. Psycholinguistic approaches to the study of communication. In L. L. Arm (Ed.), *Journeys in science: Small steps—giant strides.* Albuquerque, N. M.: University of New Mexico Press, 1967.

Miller, G. A. A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology,* 1969, **6**, 161–191.

Morris, C. W. *Signs, language, and behavior.* New York: Prentice-Hall, 1946.

Ogden, C. K., & Richards, I. A. *The meaning of meaning.* New York: Harcourt, Brace, & World, 1923.

Olson, D. R. Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review,* 1970, **77**, 257–273.

Osgood, C. E. Motivational dynamics of language behavior. *Nebraska Symposium on Motivation,* 1957, **5**.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review,* 1969, **76**, 241–263.

Perfetti, C. A. Minimal contrast in artificial word associations. *Psychonomic Science,* 1968, **10**, 229–230.

Perfetti, C. A. A study of denotative similarity with restricted word associations. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 778–795.

Perfetti, C. A., & Goodman, D. Semantic constraint on the decoding of ambiguous words. *Journal of Experimental Psychology,* 1970, **86**, 420–427.

Pollio, H. R., Deitchman, R., & Richards, S. Law of contrast and oppositional word associates. *Journal of Experimental Psychology,* 1969, **79**, 203–212.

Postal, P. M. On the surface verb "remind." *Linguistic Inquiry,* 1970, **1**, 37–122.

Quarterman, C. J., & Riegel, K. F. Age differences in the identification of concepts of the natural language. *Journal of Experimental Child Psychology,* 1968, **6**, 501–509.

Quillian, M. R. Semantic memory. In M. Minskey

(Ed.), *Semantic information processing*. Cambridge, Mass.: M. I. T. Press, 1968.

QUILLIAN, M. R. The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM,* 1969, **12,** 459–476.

REICHENBACH, H. *Elements of symbolic logic*. New York: Macmillan, 1947.

RIEGEL, K. F. Some theoretical considerations of bilingual development. *Psychological Bulletin,* 1968, **70,** 647–670.

RIEGEL, K. F. The language acquisition process: A reinterpretation of selected research findings. In L. R. Goulet & P. B. Baltes (Eds.), *Theory and research in life-span developmental psychology*. New York: Academic Press, 1970.

RIEGEL, K. F., & RIEGEL, R. M. An investigation into denotative aspects of word meaning. *Language and Speech,* 1963, **6,** 5–21.

RIEGEL, K. F., RIEGEL, R. M., QUARTERMAN, C. J., & SMITH, H. D. Developmental differences in word meaning and semantic structure. *Human Development,* 1968, **11,** 92–106.

ROMMETVEIT, R. *Words, meanings, and messages*. New York: Academic Press, 1968.

ROMNEY, A. K., & D'ANDRADE, G. G. Cognitive aspects of English kin terms. *American Anthropologist,* 1964, **66,** 146–170.

RUMELHART, D. E., NORMAN, D. A., & LINDSAY, P. H. On a semantic structure for long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization in memory*. New York: Academic Press, 1971.

SAPORTA, S. Linguistic structure as a factor and as a measure in word association. In J. J. Jenkins (Ed.), *Associative processes in verbal behavior.* Minneapolis: Department of Psychology, University of Minnesota, 1959.

SCHAEFFER, B., & WALLACE, R. The comparison of word meanings. *Journal of Experimental Psychology,* 1970, **86,** 144–152.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychological Review,* 1962, **69,** 344–354.

TULVING, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

UNDERWOOD, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology,* 1965, **70,** 122–129.

VACHEK, J. *A Prague school reader in linguistics.* Bloomington, Ind.: Indiana University Press, 1964.

WALLACE, A. F. C., & ATKINS, J. The meaning of kinship terms. *American Anthropologist,* 1960, **62,** 58–79.

WEINREICH, U. Explorations in semantic theory. In T. Sebeok (Ed.), *Current trends in linguistics.* Vol. III. The Hague: Mouton, 1966.

ZIVIAN, M. T., & RIEGEL, K. F. Word identification as a function of semantic cues and associative frequency. *Journal of Experimental Psychology,* 1969, **79,** 336–341.

(Received March 16, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 4, 260–261

# A NEGLECTED PARAMETER:
## SOME COMMENTS ON "A TABLE FOR THE CALCULATION OF $d'$ AND $\beta$"[1]

L. H. THEODOR[2]

*York University, Ontario, Canada*

Hochhaus (1972) presented a method of calculating $d'$ that is of limited generality due to its omission of the parameter $\sigma_s/\sigma_n$. A general equation for calculating $d'$ is presented, and the consequences of failing to consider $\sigma_s/\sigma_n$ are discussed in terms of some artificial data.

In a recent issue of this journal, Hochhaus (1972) presented tables for the calculation of the sensitivity and response bias parameters of signal detection theory, $d'$ and $\beta$. Hochhaus failed to point out that the method of calculating $d'$ that he presented yields a value of $d'$ that is independent of $\beta$ if, and only if, the assumption of *equal* variance in the underlying Gaussian sensory distributions is met. This paper presents the more general method of estimating $d'$, of which Hochhaus' method is a special case.

If we define $d'$ in the usual manner, $d' = (\mu_s - \mu_n)/\sigma_n$, then the appropriate equation for calculating $d'$ is:

$$d' = \sigma_s/\sigma_n \; Z\;(\text{HR}) - Z\;(\text{FAR}) \qquad [1]$$

where HR = hit rate and FAR = false alarm rate. Since $\sigma_n$ may always be set equal to unity, the term $\sigma_s/\sigma_n$ represents the scaling factor necessary to convert $Z$ (HR) into $\sigma_n$ units. In practice, $\sigma_s/\sigma_n$ is equal to the reciprocal of the slope of the receiver operating characteristic, plotted in normal deviate coordinates (Green & Swets, 1966). The equation for calculating $d'$ that Hochhaus has presented is for the special case when $\sigma_s = \sigma_n$.

In many experiments, particularly in vision, the equal variance assumption is grossly untenable (cf. Green & Swets, 1966). The consequence of failing to take $\sigma_s/\sigma_n$ into account are illustrated in Table 1 and Figure 1, which present some artificial, though not unlikely, data from a single observer under three conditions of a hypothetical experiment. As seen in Figure 1, these data form a linear operating characteristic with a slope of .5 when plotted in normal deviate coordinates, characterized by an $X$ intercept of $-1.5$. Hochhaus' equation would lead us to assume that the performance under each condition arose from three different operating characteristics, each with unit slope and intercepts of $-1.5$, $-1.0$, and $-.5$, respectively, as shown by the dashed functions in Figure 1. To know which of these cases is true would require knowledge of the slope of the normal deviate operating characteristic under each of the three conditions of the experiment. Such knowledge may be obtained by the use of the confidence-rat-

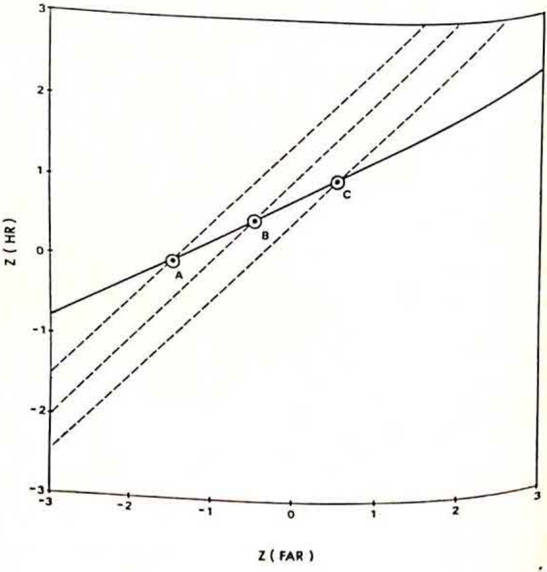


FIG. 1. Data from Table 1 plotted in normal deviate coordinates (HR = hit rate; FAR = false alarm rate).

---

[1] This research was supported by Grant APA-300 from the National Research Council of Canada.

[2] Requests for reprints should be sent to L. H. Theodor, Department of Psychology, York University, 4700 Keele Street, Downsview 463, Ontario, Canada.

TABLE 1

ARTIFICIAL DATA FOR A SINGLE SUBJECT UNDER THREE CONDITIONS OF AN EXPERIMENT WITH $d'$ CALCULATED UNDER THE ASSUMPTION THAT $\sigma_s/\sigma_n = 1$ AND $\sigma_s/\sigma_n = 2$

| Condition | HR | FAR | $Z$ (HR) | $Z$ (FAR) | $d'$ | |
|---|---|---|---|---|---|---|
| | | | | | $\sigma_s/\sigma_n = 1$ | $\sigma_s/\sigma_n = 2$ |
| A | .5000 | .0668 | 0 | −1.5 | 1.5 | 1.5 |
| B | .6915 | .3085 | .5 | −.5 | 1.0 | 1.5 |
| C | .8413 | .6915 | 1.0 | .5 | .5 | 1.5 |

Note.—HR = hit rate; FAR = false alarm rate.

ing procedure and a computer program such as that of Dorfman and Alf (1969) for fitting receiver operating characteristic functions.

Two major points emerge from the above analysis: First, $d'$ is independent of response bias only when $\sigma_s/\sigma_n$ has been taken into account; and second, it is necessary to know $\sigma_s/\sigma_n$ in order to properly calculate $d'$.

## REFERENCES

DORFMAN, D., & ALF, E., JR. Maximum-likelihood estimation of parameters of signal-detection theory—rating method data. *Journal of Mathematical Psychology,* 1969, 6, 487–496.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

HOCHHAUS, L. A table for the calculation of $d'$ and $\beta$. *Psychological Bulletin,* 1972, 77, 375–376.

(Early publication received May 23, 1972)

*Psychological Bulletin*
1972, Vol. 78, No. 4, 262–265

# TEST OF NORMALITY AGAINST SKEWED ALTERNATIVES

RALPH B. D'AGOSTINO[1]
*Boston University*

EDWARD E. CURETON
*University of Tennessee*

A useful test of normality is the test based on $g_1$, the third standardized sample moment. The test is computationally simple and has good sensitivity to nonnormality due to skewness. It is asymptotically the most powerful test of normality against deviations from normality due solely to skewness. This test is described, and two simple approximations to its null distribution are given. The first is a normal approximation valid for large samples, and the second is a $t$ transformation valid for sample sizes as small as eight. Some general comments concerning the use of tests of normality are also presented.

In an article by D'Agostino (1970a) Geary's test was suggested to research psychologists as a test of normality. As mentioned there and illustrated by D'Agostino and Rosman (1969), this test possesses good sensitivity to deviations from normality, especially if the deviations are due to kurtosis. If, however, the deviations are due to skewness (asymmetry), a test usually more powerful than Geary's is the test based on $g_1$, the third standardized sample moment (see D'Agostino & Rosman, in press; Shapiro, Wilk, & Chen, 1968). (Some authors refer to $g_1$ as $\sqrt{b_1}$.) Asymptotically (i.e., for large samples) it is the most powerful test of normality against deviations from normality due solely to skewness (Geary, 1947). When compared to standard tests such as the chi-square and Kolmogorov-Smirnov, this test possesses all the excellent features that Geary's test does. It requires little computation, almost no manipulation of the sample (such as ordering of the observations or arranging then in a frequency table), no knowledge of the population mean and standard deviation, and it applies to samples of all sizes.

The $g_1$ test is illustrated by testing the null hypothesis of normality for the fictitious data given in Table 1. The test statistic $g_1$ is the third standardized sample moment (i.e., the ratio of the third central sample moment to the sample variance raised to the 1.5th power). In symbols:

$$g_1 = \frac{\dfrac{\Sigma(X-\bar{X})^3}{N}}{\left[\dfrac{\Sigma(X-\bar{X})^2}{N}\right]^{\frac{3}{2}}} = \frac{\sqrt{N}\Sigma(X-\bar{X})^3}{[\Sigma(X-\bar{X})^2]^{\frac{3}{2}}} \quad [1]$$

where $X$ represents an observation, $\bar{X}$ the sample mean, and $N$ the sample size. A computational formula is readily available by expanding the numerator and denominator of Equation 1. It is:

$$g_1 = \frac{N^2\Sigma X^3 - 3N\Sigma X\Sigma X^2 + 2(\Sigma X)^3}{[N\Sigma X^2 - (\Sigma X)^2]^{\frac{3}{2}}}. \quad [2]$$

The data in Table 1 yield $g_1 = -.5242$.

The next step, after the computation of $g_1$, is to decide whether or not the null hypothesis of normality should be accepted or rejected at a selected level of significance. If the null hypothesis of normality is true, the expected value of $g_1$ is zero. Negative (positive) skewness produces on the average negative (positive) values. One usually rejects the null hypothesis for large deviations in absolute value of $g_1$ from zero (i.e., a two-sided test). For certain significance levels (.02 and .10) a table of critical values exists in Pearson and Hartley's (1966) *Biometrika Tables for Statisticians* (Volume 1). This table is for samples of size $N = 25(5)50(10)100(25)200(50)500$. No entries exist for $N = 9$, which is needed for our example in Table 1. To handle this and other sample sizes we now present and investigate two approximations for $g_1$.

If the null hypothesis of normality is true, then $g_1$ is asymptotically normal with mean zero and variance $6(N-2)/[(N+1)(N+3)]$.

---

[1] Also at the Educational Research Corporation.

Requests for reprints should be sent to Ralph B. D'Agostino, Mathematics Department, Boston University, Charles River Campus, Boston, Massachusetts 02215.

TABLE 1

DATA ILLUSTRATING THE $g_1$ TEST

| $X$ | $(X - \bar{X})^2$ | $(X - \bar{X})^3$ | $X^2$ | $X^3$ |
|---|---|---|---|---|
| 8 | 36 | −216 | 64 | 512 |
| 9 | 25 | −125 | 81 | 729 |
| 13 | 1 | −1 | 169 | 2197 |
| 14 | 0 | 0 | 196 | 2744 |
| 15 | 1 | 1 | 225 | 3375 |
| 15 | 1 | 1 | 225 | 3375 |
| 16 | 4 | 8 | 256 | 4096 |
| 17 | 9 | 27 | 289 | 4913 |
| 19 | 25 | 125 | 361 | 6859 |
| Total 126 | 102 | −180 | 1866 | 28800 |

Note.—$N = 9$; $\bar{X} = 14$. Computation using Equation 1: $\Sigma(X - \bar{X})^3 = -180$; $N\Sigma(X - \bar{X})^3 = -540$, $\Sigma(X - \bar{X})^2 = 102$

$$g_1 = \frac{-540}{[102]^{3/2}} = \frac{-540}{1030.2} = -.5242.$$

Computation using Equation 2:

$$\begin{aligned} N^2\Sigma X^3 - 3N\Sigma X\Sigma X^2 + 2(\Sigma X)^3 &= 9^2(28{,}000) - 3(9)(126)(1{,}866) + 2(126)^3 \\ &= 14{,}580. \\ N^2\Sigma X^2 - (\Sigma X)^2 &= 9(1{,}866) - 126)^2 \\ &= 918 \end{aligned}$$

$$g_1 = \frac{-14{,}580}{[918]^{3/2}} = \frac{-14{,}580}{27{,}815.4} = -.5242.$$

Thus,

$$Y = g_1\sqrt{\frac{(N+1)(N+3)}{6(N-2)}} \qquad [3]$$

is asymptotically a standard normal variable (i.e., mean zero and unit variance). One can use standard normal tables for the test if $N$ is large. For example, for a two-tailed test one rejects normality at the 10% level of significance if $|Y| \geq 1.645$. The usual "rule of thumb" is that the normal approximation of Equation 3 is valid for $N \geq 150$ and for those sample sizes $g_1\sqrt{N/6}$ is an adequate approximation to Equation 3 (Snedecor & Cochran, 1967, p. 86).

Table 2 contains the results of a simulation study of the approximation given by Equation 3 where 10,000 random normal samples were generated for samples of size $N = 8$, 10, 15, 20, 25, 50, 100, 125, and 150. For each sample the $Y$ statistic of Equation 3 was computed and classified into the nested intervals formed by the critical values corresponding to assuming that $Y$ is a standard normal variable for two-tailed tests with levels .01, .02, .05, .10, and .20. The entries are the Monte Carlo probabilities (relative frequencies) for these nested intervals. Ideally, if the approximation is correct and there is no sampling error, they would be .010, .020, .050, .100, and .200. Only for $N = 150$ are the differences between the Monte Carlo probabilities and the desired levels of significance within sampling error as measured by the chi-square test. However, if one restricts himself to tests with levels .05, .10, and .20, based on these simulation results, the approximation is probably adequate for most practical purposes (usually less than a 10% error). For our example in Table 1, $Y = -.8864$ and with, say, a two-tailed test at the 10% level of significance, the decision is to accept the hypothesis of normality since $|Y| = .8864 < 1.645$.

The second approximation, which is valid for $N \geq 8$, arises from approximating the null distribution of $g_1$ by a Pearson-type VII curve (i.e., a $t$ distribution) (Williams, 1935). It is as follows: Assume the hypothesis of normality is true and $N \geq 8$. Let

$$\beta_2 = \frac{3(N^2 + 27N - 70)(N+1)(N+3)}{(N-2)(N+5)(N+7)(N+9)} \qquad [4]$$

TABLE 2

MONTE CARLO PROBABILITIES FOR VARIOUS LEVELS OF SIGNIFICANCE USING APPROXIMATION $Y$ OF EQUATION 3

| Desired level of significance[a] | $N$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 8 | 10 | 15 | 20 | 25 | 50 | 100 | 125 | 150 |
| .01 | .013 | .015 | .015 | .014 | .016 | .013 | .012 | .014 | .010 |
| .02 | .024 | .026 | .026 | .025 | .026 | .022 | .024 | .022 | .019 |
| .05 | .056 | .058 | .053 | .051 | .054 | .054 | .050 | .053 | .048 |
| .10 | .105 | .104 | .096 | .094 | .098 | .100 | .101 | .098 | .094 |
| .20 | .199 | .197 | .188 | .181 | .187 | .189 | .195 | .189 | .192 |

Note.—10,000 samples per sample size.

[a] Two-tailed test.

and

$$v = \frac{4\beta_2 - 6}{\beta_2 - 3}, \qquad [5]$$

then

$$t = g_1\sqrt{\left(\frac{v}{v-2}\right)\left(\frac{(N+1)(N+3)}{6(N-2)}\right)} \qquad [6]$$

is distributed approximately as a $t$ variable with $v$ degrees of freedom.

The parameter $v = 48.20020$, for $N = 8$, dips to a minimum of $v = 14.31359$ for $N = 22$, and then steadily rises ($v = 17.26094, 25.09627$, and $41.54076$ for $N = 50$, 100, and 200, respectively). While $v$ is apparently never an integer, interpolation between the entries in the usual $t$ tables should be adequate for most purposes. For the needed range of $v$s the change in the $t$ values for one integer $v$ value to the next is usually at most only one unit in the second decimal place.

We have two approaches for judging the accuracy of this $t$ transformation. First, for $N \geq 25$, we can use the already mentioned *Biometrika* (Pearson & Hartley, 1966) tables. Our approach is to compare the critical $g_1$ values of that table with those obtained by using the inverse of the transformation given from Equations 4 to 6. For $N > 150$, no differences exist. For $25 \leq N \leq 150$, the greatest discrepancy is only four units in the third decimal place, a relative error never exceeding .01 standard deviation units. For $N < 25$, we generated 20,000 random normal samples of size $N = 8$, 10, 15, and 20. For each sample, the $g_1$ statistic was computed and classified into the nested intervals formed by the two-sided tests with level of significance .01, .02, .05, .10, and .20. In Table 3 the results are presented. The entries are the Monte Carlo probabilities. Without fail the Monte Carlo results agree with the desired levels to at least two decimal places. For most practical applications this agreement is more than adequate.

TABLE 3

MONTE CARLO PROBABILITIES FOR VARIOUS LEVELS OF SIGNIFICANCE USING $t$ APPROXIMATION

| Desired level of significance[a] | N | | | |
|---|---|---|---|---|
| | 8 | 10 | 15 | 20 |
| .01 | .010 | .011 | .010 | .011 |
| .02 | .022 | .022 | .022 | .022 |
| .05 | .053 | .054 | .049 | .052 |
| .10 | .100 | .102 | .099 | .010 |
| .20 | .196 | .198 | .197 | .198 |

Note.—20,000 samples per sample size.
[a] Two-tailed test.

For the data of Table 1, we have with $N = 9$ the $\beta_2$ of Equation 4 equal to 3.2398 and the $v$ of Equation 5 equal to 29.0213. Using the $t$ distribution with 29 degrees of freedom and a level of significance of .10, the critical values are ±1.699. The observed $t$ found by using Equation 6 yields $-.9183$. Thus our decision is to accept the hypothesis of normality.

Skewness due to a lengthening of one tail of the distribution as compared with the other almost invariably implies nonmesokurtosis. Hence, Geary's test will tend to reject normality against skew alternatives of this commonest type whenever $g_1$ does, but with somewhat lower power, and it also rejects normality against symmetric platykurtic and leptokurtic alternatives. On the other hand, $g_1$ has zero power against these latter alternatives.

There is at least one type of asymmetric nonnormal alternative distribution against which neither $g_1$ nor the Geary test has any appreciable power. This distribution is formed from a normal distribution by moving the mode upward, say by $\frac{1}{4}\sigma$ to $\frac{1}{2}\sigma$ without increasing or decreasing its height, and then shifting some of the frequency from the intermediate region (*not* the tail) above the mode to the intermediate region below the mode, resulting in a bulge on the side of the curve below the mode, mainly in the region from the mean to about $-1\frac{1}{2}\sigma$ or $-2\sigma$. A distribution of this type *can* be exactly "symmetric" by the $g_1$ test and exactly mesokurtic by the Geary test. An example of such a distribution is that of test intelligence in the World War II army (Staff, Personnel Research Section, 1947). For alternative distributions of this type, recourse must still be had to the chi-square and Kolmogorov-Smirnov tests despite their low powers, and large samples are required to reject the hypothesis of normality even though it is easily observed in smaller samples when a histogram or frequency polygon is prepared.

Figure 1 shows an artificial distribution of this type, with $N = 200$. For this distribution, the mean is 12.320, the median is 12.447, and

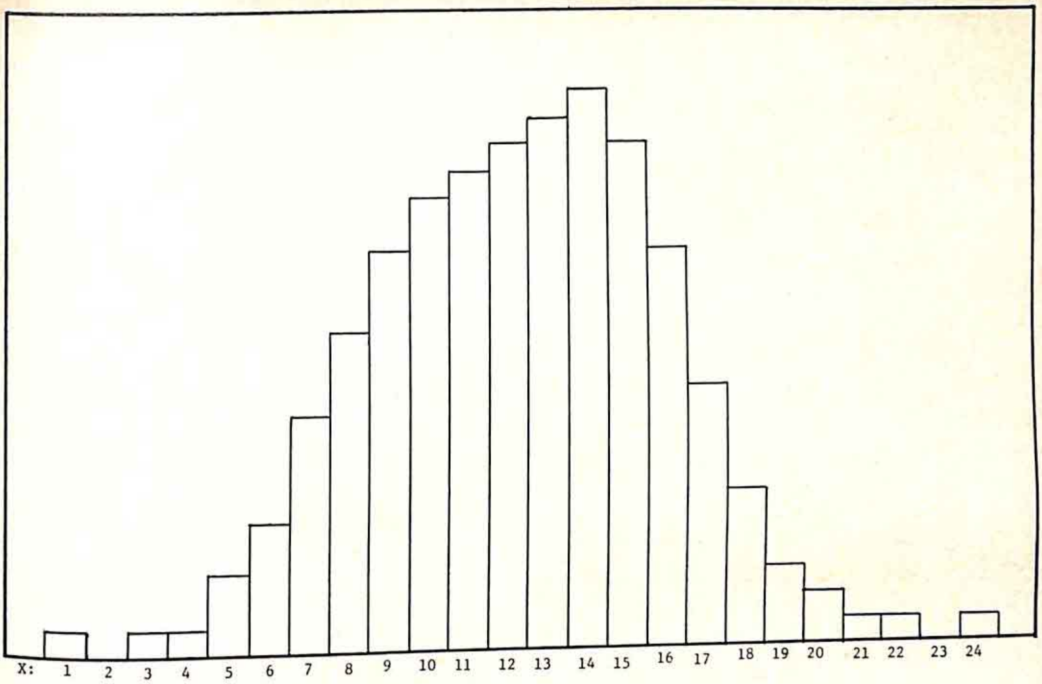


FIG. 1. Artifical frequency distribution.

the crude mode is 21. Note that the rough rule, (mode-mean) = 3 (median-mean) breaks down completely for this distribution. The standard deviation is 3.745, $g_1 = -.0098$, and Geary's $a = .8065$. Even if every frequency were multiplied by 5, so that $N = 1000$, neither $g_1$ nor $a$ would be significant at the .05 level.

## REFERENCES

D'AGOSTINO, R. B. Simple compact portable test of normality: Geary's test revisited. *Psychological Bulletin*, 1970, **74**, 138–140. (a)

D'AGOSTINO, R. B., & ROSMAN, B. Reconsideration of Geary's test of normality. *Journal of the American Statistical Association*, in press.

GEARY, R. C. Testing for normality. *Biometrika*, 1947, **34**, 209–242.

PEARSON, E. S., & HARTLEY, H. O. *Biometrika tables for statisticians*. Vol. I. (3rd ed.) Cambridge: Cambridge University Press, 1966.

SHAPIRO, S. S., WILK, M. B., & CHEN, H. J. A comparative study of various tests of normality. *Journal of the American Statistical Association*, 1968, **63**, 1343–1372.

SNEDECOR, G. W., & COCHRAN, W. G. *Statistical methods*. (6th ed.) Ames, Iowa: Iowa State University Press, 1967.

STAFF, PERSONNEL RESEARCH SECTION, Army General Office. The Army General Classification Test, with special reference to the construction and standardization of Forms 1a and 1b. *Journal of Educational Psychology*, 1947, **38**, 385–420.

WILLIAMS, P. Note on the sampling distribution of $\sqrt{b_1}$, where the population is normal. *Biometrika*, 1935, **27**, 269–271.

(Received January 20, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 4, 266–278

# PERCEPTION AND DISCRIMINATION AS A FUNCTION OF STIMULUS ORIENTATION:

## THE "OBLIQUE EFFECT" IN MAN AND ANIMALS

STUART APPELLE [1]

*State University of New York at Brockport*

Performance for a large variety of perceptual tasks is superior for stimuli aligned in horizontal or vertical orientations, as compared to stimuli in oblique orientations. This phenomenon appears in the human adult and child, and throughout the animal kingdom. Neurophysiological mechanisms for orientation analysis have been found in the higher visual pathways of many animals, and the suggestive evidence is compelling that these mechanisms underly the orientation preferences reported behaviorally. This paper reviews both the behavioral and neurophysiological studies of orientation preferences, and suggests additional methods for determining the cause of these effects.

Stimulus orientation has received increased attention in the past decade from neurophysiological studies of visual pathways. These studies have located mechanisms for orientation analysis in single cells of mammalian visual systems. The study of stimulus orientation, however, has a long history in the behavioral and psychophysical literature. A persistent feature of these latter studies has been a small but consistent superiority in performance when visual stimuli are horizontal or vertical, as opposed to oblique. (For convenience, this phenomenon is subsequently referred to as the *oblique effect*.) The origin of the oblique effect has not been precisely determined, but a relationship between the behavioral results and the neurophysiological data is suggestive.

This paper deals with the oblique effect in two parts. First, it reviews the behavioral and psychophysical studies of orientation preferences. Second, it discusses the neurophysiological substrates of orientation perception. The first section helps define the oblique effect by elaborating on the many forms in which it appears. The second section amplifies the first by evaluating the neurophysiological findings in respect to these orientation preferences.

## BEHAVIORAL STUDIES

### *Man*

### *Acuity*

Studies of visual acuity in man were among the first investigations to uncover preferences for vertical and horizontal stimuli. Emsley (1925) noticed acuity differences among subjects asked to resolve line gratings. Maximal acuity occurred when the gratings were in horizontal or vertical orientations. Visibility was poorest for stimuli 45 degrees or 110–140 degrees. Emsley initially attributed these results to astigmatism, but when lenses were used to correct for this disorder the effect remained. Emsley called this phenomenon "residual astigmatism."

Later studies (Hamblin & Winser, 1927; Higgins & Stultz, 1948; Leibowitz, 1953; Shlaer, 1937) confirmed these results for grating acuity, and also added a new dimension to the problem. It was discovered (Higgins & Stultz, 1948) that by tilting the head, so as to align the oblique stimuli with the retinal field, diagonal visibility improved. It seemed that orientation of the image was more important than orientation of the object, but later results have been more ambiguous on this point. For example, Luria (1963) determined acuity (in terms of pattern density) for checkerboard patterns while subjects were in upright or head-tilted conditions. Although vertical targets were significantly more visible than oblique patterns, this difference decreased with amount of head tilt. Attneave and Olson (1967) concluded from one experiment that orientation of the stimulus was most important, while they concluded the reverse from a second experiment (Attneave & Olson, 1970). Meanwhile, Attneave & Reid (1968) have suggested that perceived orientation, rather than actual orientation either in space or on the retina, is the critical factor. The ambiguity in these studies is further complicated by neurophysiological evidence from the cat, which suggests that cortical cells may compensate for head tilt by altering their response (Horn & Hill, 1969).

---

[1] The author expresses appreciation to R. D. Walk, E. Abravanel, and L. Rothblat for their valuable ideas and comments.

Requests for reprints should be sent to Stuart Appelle, Department of Psychology, State University of New York, College at Brockport, Brockport, New York 14420.

The difficulty with obliques on acuity tests was later found for a Landolt C test object (Hartridge, 1947), and Leibowitz (1955) demonstrated that the problem occurred in vernier acuity as well. Ogilvie and Taylor (1958, 1959) measured the visibility of very fine wires and found that oblique wires had to be twice as wide as vertical or horizontal wires to be seen.

### *Adjustment and Assessment of Stimulus Orientation*

Perhaps the earliest discovery of an oblique effect was made by Jastrow (1893). Subjects required to reproduce visually presented lines, or to set other lines to some specified orientation, showed marked superiority with horizontal and vertical stimuli. More recent investigations have confirmed Jastrow's results. Smith (1962), for example, required observers to estimate the headings (radial position) of radar trails. Minimal errors were found at the horizontal and vertical axes. Discrepancies from these axes seemed to be perceptually exaggerated. Bouma and Andriessen (1968) determined perceived orientation by requiring subjects to adjust a dot to the perceived extension of a line segment. They tended to set the dots toward the nearest horizontal or vertical coordinate. Oblique lines seemed more horizontal or vertical than they were (by 3–10 degrees). Accuracy was lower for the oblique orientations and standard deviation was 2.5 times as great as for horizontal and vertical lines.

Sulzer and Zener (1953) determined the effect of orientation on the perception of parallelism. Two black lines were presented on a white ground. One line could be rotated by the subject to bring it parallel with the second reference line. Variability of the settings was minimal with horizontal reference lines, poorer with vertical lines, and maximal with obliques. These results were replicated by Onley and Volkman (1958) and extended by Rochlin (1955). Andrews (1965, 1967a, 1967b) has pursued a similar analysis. Subjects were instructed to set a short bright line to an orientation parallel to a nearby standard or to make a forced-choice decision as to the lines' parallelism. When the test line was in a horizontal or vertical position, variability was at a minimum and acuity was best. Both Andrews and Bouma and Andriessen interpreted their experiments in terms of neural units responsive to particular orientations. Bouma and Andriessen (1968) have suggested a denser population of horizontal and vertical units. Andrews (1967a, 1967b) has suggested these units have a greater sensitivity.

These results are not restricted to straight line test objects. Leibowitz, Myers, and Grant (1955) obtained a measure of localization accuracy for circular test fields, 35 minutes in arc. The lights were presented in different radial positions, 4.2 degrees from a central fixation point. Localization accuracy was found to be poorest around the oblique positions and best at vertical and horizontal coordinates. In another experiment (Burns, Mandl, Pritchard, & Webb, 1969) observers were asked to give subjective impressions of target form and shape to a point source of light. Targets were generally reported as straight lines, while the slope was typically seen as near to horizontal regardless of stimulus position. Burns et al. suggest that some preferential treatment of horizontal stimuli in the visual system of man is implied in these data.

### *Other Studies on Object Orientation*

Taylor (1963) has taken note of some studies to point out the difficulties with oblique positions; both the detection of stimuli and the assessment of line orientation are probably two aspects of the same problem. Meanwhile, other investigators have reported oblique effects in still other situations. For example, Craig and Lichtenstein (1953) had subjects report perceptual changes to a line during extended periods of fixation. Minimal disappearance occurred when the test line was along the vertical or horizontal axis, while maximal disappearance centered around a 45-degree oblique position. Similarly, McFarland (1968) reported that right-angle stabilized images disappear with a much lower frequency than acute and obtuse angles or single horizontal lines.

Foley (1962) was concerned with the contribution of differential summation to the oblique effect. He hypothesized that area and luminance functions should be less for oblique stimuli in a critical flicker frequency test. The functions obtained, however, were parallel for a 60-degree oblique and a horizontal test line, indicating that observed differences with oblique and horizontal stimuli cannot be attributed to differential summation. Critical flicker frequency, however, was always lower with the oblique stimulus, a result that implies a lower effective luminance for stimuli in this orientation.

Several studies have examined the role of stimulus orientation in perceptual grouping. Beck (1966a, 1966b, 1967) presented stimuli composed of patterns of figure elements (short lines) arranged in different orientations. The figure elements in any one orientation composed a subdivision of the total figure. Subjects were required to segregate two overlapping stimulus fields, or divide three-boundary figures into two regions according to the division most naturally perceived in the overall figure. The results of these studies suggested that the most facilitating aspect of perceptual grouping was the change from horizontal or vertical orientations to diagonal orientations in the component lines of the figure elements.

Optical illusions are also subject to the oblique effect. The Poggendorf illusion consists of an oblique line that appears to pass behind a rectangle. This objectively straight line often looks broken or displaced. Green and Hoyle (1964) found that placing the oblique line in a horizontal orientation greatly reduces the illusion. Leibowitz and Toffey (1966) found that the illusion was reduced by one-half when the line was horizontal or vertical. The Zöllner illusion consists of long parallel lines intersected by a series of 45-degree slashes. The parallel lines often appear to be slanted in respect to each other. This illusion is minimal when the lines are in a horizontal or vertical position and maximal for 45- and 135-degree obliques (Judd & Courten, 1905; Wallace, 1964). These studies suggest that rotating the distorted lines into vertical or horizontal positions greatly reduces the strength of the illusions.

### *Children*

Children as well as adults are susceptible to the oblique effect. Gibson, Gibson, Pick, and Osser (1962) tested 4–8-year-olds on visual discrimination of letterlike forms. Standard forms were matched against comparisons rotated 45, 90, or 180 degrees. The greatest errors occurred with the 45-degree transformations. Rudel and Teuber (1963) tested $3\frac{1}{2}$–$8\frac{1}{2}$-year-old children with vertical, horizontal, and oblique lines and U shapes. Stimuli were simultaneously presented in pairs, and the children were asked to guess which member of the pair was "correct." While virtually no children failed to learn the horizontal–vertical discrimination, almost all the 3–5-year-olds and one-third of the $8\frac{1}{2}$-year-olds failed to discriminate left and right obliques. Also, successive presentations proved even more difficult to master, a fact that suggested a memorial component to the problem.

Over (1967), in fact, has suggested that the oblique effect may be due to memory variables per se rather than any visual mechanisms. This hypothesis was tested by Over and Over (1967) with $3\frac{1}{2}$–$6\frac{1}{2}$-year-old children. Two procedures were used: One (recognition task) required the children to remember an oblique stimulus as an oblique from trial to trial, and the other (detection task) only required that the correct stimulus be detected from a two-comparison array. The experimenters found that most children could distinguish between obliques with the detection task but that many failed to discriminate obliques with the recognition task. All subjects, however, could make a horizontal–vertical discrimination. Although this result suggests the role of memory, when time intervals of 15–90 seconds separated presentation of standard and comparison in a third task, many previously unsuccessful subjects were now successful at all time delays. Over and Over (1967) suggested that some factors beyond memory must be involved in the recognition of mirror-image obliques.

Bryant (1969) tested 5–7-year-olds on simultaneous and successive matching tasks and found successive discrimination of mirror-image obliques to be most difficult. Bryant then went on to test difficulty on non-mirror-image obliques. Since difficulty in discriminating between mirror-image stimuli among children has been demonstrated for stimuli other than oblique lines (Huttenlocher, 1967; Rudel & Teuber, 1963; Sekular & Rosenblith, 1964) performance with 45-degree versus 135-degree objects could be ascribed to either the mirror image or oblique characteristic of the discrimination. Comparison of two obliques led to chance-level discrimination whether they were mirror images or not. Bryant concluded that oblique lines are seen as oblique but that the particular direction of slope is not encoded in memory.

## *Animals*

Preferences for horizontal and vertical stimuli appear in many different organisms. These preferences are investigated by training animals to differentially respond to two or more stimuli to obtain reward (usually food). Rate of learning can thereby be used as a measure of ability to discriminate between objects in different orientations.

### *Octopus*

Sutherland (1957) trained octopuses to respond to various rectangular bar shapes. The animal was rewarded with a crab when it attacked a positive rectangle and was shocked if it attacked a negative rectangle. It learned to respond to horizontal and vertical bars with about 90% accuracy. With a left–right oblique discrimination, however, there was no evidence of learning after 100 trials. Sutherland (1958) later tested octopuses on horizontal versus oblique and vertical versus oblique discriminations. Both were considerably more accurate than an oblique–oblique discrimination and much less accurate than a vertical–horizontal discrimination. Wells (1960) has shown that orientation discrimination in the octopus is dependent on the position of the retinal image. In the normal animal, the eyes rotate so that they are always horizontal to the environment regardless of the animal's position. When the octopus' statocysts are removed, the eyes stay horizontal in relation to the head rather than in relation to the gravitational (environmental) axes. Octopuses are incapable of discriminating orientation after the statocysts are removed.

### *Goldfish*

Mackintosh and Sutherland (1963) have shown that goldfish have difficulty learning to discriminate objects in oblique orientations. The fish were trained to take food from rewarded rectangular shapes. One group of fish was trained to discriminate between horizontal and vertical rectangles, a second group between left and right oblique rectangles. The second group required twice as many trials to learn the oblique discrimination.

### *Pigeon*

Zeigler and Schmerler (1965) presented pigeons with white rectangles on black backgrounds in vertical, horizontal, 60-degree, and 120-degree orientations. These stimuli were projected on keys to which the pigeons responded for reward. A 15-second fixed-interval reinforcement schedule was used for training. Both simultaneous and successive discriminations were studied. Discrimination, as measured by differential rate of responding, was found to be essentially the same for vertical–horizontal, vertical (or horizontal)–oblique, and oblique–oblique discriminations. Although oblique stimuli do not seem to present a problem to pigeons in a discrimination task, Mello (1965, 1966a, 1966b, 1967) has shown that pigeons monocularly trained to respond to an oblique line will respond maximally to its mirror image when the untrained eye is tested. It has been argued (Beale & Corballis, 1968), however, that this "paradoxical transfer" phenomenon may be a procedural artifact. This issue has not yet been satisfactorily resolved.

### Rabbit

Rabbits have been trained to discriminate either rectangular bars (Van Hof, 1965) or rows of dots (Van Hof, 1969) in different orientations. These studies found that both horizontal–vertical and oblique–oblique discriminations could be performed with close to 100% accuracy. Rate of acquisition, however, is slower for oblique discriminations (Van Hof, 1970a, 1970b). Van Hof and Wiersma (1967) determined the minimum angular difference that could be detected around a horizontal, vertical, and 45-degree oblique axis. Threshold discrimination was 5–10 degrees around the vertical and horizontal axes, and 10–15 degrees around the 45-degree oblique axis. These studies suggest that although rabbits can learn oblique discriminations, obliques present more difficulty than vertical–horizontal discriminations.

### Rat and Squirrel

An early and comprehensive study on discriminative abilities in rats by Lashley (1938) revealed that rats learned to discriminate more rapidly between vertical and horizontal striated patterns than between 45- and 135-degree striations. Koronakos and Arnold (1957) studied the formation of learning sets in rats using eight different oddity problems. In one problem a vertically striped door had to be selected from four other horizontally striped doors. Another problem required a left-oblique-lined door to be selected from four other right-oblique-lined doors. This second problem proved to be the more difficult to master.

Dodwell (1970) has recently shown that rats' difficulty with oblique discriminations persists, regardless of the type of testing apparatus used. In addition, Dodwell compared the discrimination of rats with squirrels. Like rats, squirrels had more difficulty learning to discriminate between a diagonal and vertical than a horizontal and vertical. A diagonal-horizontal discrimination could not be learned at all.

### Cat

Sutherland (1963a) trained cats to discriminate between horizontal and vertical, or 45-degree and 135-degree oblique rectangles. The cats learned to discriminate either problem in an average of about 11 days. Performance on the horizontal–vertical problem was 75% correct; performance was 78% correct on the mirror-image obliques. Warren (1969) replicated these results on another apparatus, while eliminating the possibility of influence from contextual cues.

The problem presented to Sutherland's (1963a) and Warren's (1969) cats was a simultaneous discrimination. Parriss (1964) was interested in cats' performance when a successive discrimination was required. Four cats responded to rectangular shapes on two doors for food. The cats were confronted with a locked door when responding to the negative door. Training covered 20 trials per day, for five consecutive days each week. Training was terminated when the animal's performance was 90% correct over 80 consecutive trials. Two cats learned a horizontal–vertical discrimination first, then a 45–135-degree discrimination. The other cats learned in the opposite order. In the first group, the oblique discrimination took twice as many trials as the horizontal–vertical task and produced three and five times as many errors. When the oblique discrimination was the first problem learned, three times as many trials were required, and four and six times as many errors were made. It would seem that cats are no better with oblique stimuli than other animals, at least when the stimuli are presented in successive fashion.

### Chimpanzee

Chimpanzees are close to man phylogenetically and have similar visual systems. It would be interesting to know how well they can discriminate orientations, however, not many studies are available on these animals' orientation preferences. Nissen and McCulloch (1937) tested monkeys, but their main concern was the effect of different learning techniques on discrimination, and it is difficult to interpret their data in terms of the oblique effect. However, the data do suggest that the animals had more difficulty learning to discriminate between opposing diagonally striped patterns than between vertical and horizontal stripes. One animal failed to discriminate diagonal patterns after 800 trials.

### *Behavioral Studies: Summary*

Wherever stimulus orientation has been manipulated, performance has been consistently better with objects in vertical or horizontal orientations, as compared to stimuli in oblique orientations. This phenomenon has been demonstrated for the resolving power of the human eye (grating acuity, vernier acuity, Landolt C test object, and fine-line acuity), critical flicker frequency, fading of stabilized images, setting of lines to specified orientations, estimation of stimulus position, reaction times, perceptual grouping, matching and learning tasks, stimulus discrimination and generalization, and optical illusions. It has appeared in the human adult and child, and in the octopus, goldfish, pigeon, rabbit, squirrel, rat, cat, and chimpanzee. These studies clearly establish the reality of preferences for vertical and horizontal orientation. They are less clear in identifying the origin of these preferences. One approach to that problem has been to look for underlying mechanisms in the neurophysiological substrates of orientation perception. In reviewing the behavioral studies, it has occasionally been necessary to refer to these systems. The second section of this paper examines the neurological evidence in detail, and pays special attention to any possible connection to oblique effects.

## NEUROPHYSIOLOGICAL MECHANISMS OF ORIENTATION ANALYSIS

A number of attempts have been made to relate the psychophysical evidence for an oblique effect with anatomical or neurophysiological data. The approach is an encouraging one. Sutherland (1963b) has referred to

> the exciting possibility that when both neurophysiological and behavioral evidence are available for a variety of species, we may be able to correlate variations in behavior with variations in the proportions of different types of receptive fields present in the brain of different species [p. 122].

With this in mind, the present section reviews the neurophysiological studies of orientation perception.

### *Animals*

The visual systems of a number of animals have been studied, but the earliest and most exhaustive studies have been carried out on the cat. It seems advisable to discuss these investigations first. Hubel and Wiesel (1962), using perfected microelectrode techniques, were able to record electrical activity from single cells in the cat's visual cortex. Anesthetized animals were presented with bar-shaped stimuli, which were moved around the visual field while the recordings were being made. By mapping out the region of the retina that could influence the firing of particular cells, the "receptive field" of those cells could be determined. Hubel and Wiesel discovered that the cells in cat striate (Hubel & Wiesel, 1962, 1963) and nonstriate cortex (Hubel & Wiesel, 1965) responded maximally to bars or contours of particular size, position, and orientation. Departure from the optimal characteristics would diminish or eliminate that cell's response. For example, a change of 5–10 degrees in orientation could abolish a cell's response completely. Monkey cortex was found to be organized in a way quite similar to that described for the cat (Hubel & Wiesel, 1968). The pigeon retina (Maturana & Frenk, 1963), goldfish tectum (Jacobson & Gaze, 1964), and rabbit ganglion cell (Levick, 1965, 1967) also seem to be organized to discriminate orientation.

Several studies of restricted rearing have confirmed the dependence of orientation discrimination on orientation-sensitive neurons in the cat. Wiesel and Hubel (1965) and Ganz, Fitch, and Satterberg (1968) have shown that if kittens are deprived of early visual experience, cortical cells will respond with no specificity with regard to orientation. The kittens will be behaviorally blind. Ganz and Fitch (1968) reared cats with one eye shut for 1–6 months. With the experienced eye, 85% accuracy on a vertical–horizontal discrimination was reached after 100 trials. There was no learning after 2,000 trials with the nonexperienced eye. More specific demonstrations of the relationship between experience and cortical function have come from Blakemore and Cooper (1970) and Hirsch and Spinelli (1970). These investigators showed that cells in cat cortex will respond only to the particular orientations experienced during rearing.

The studies on visual deprivation, along with the microelectrode evidence from normal animals, strongly support the contention that cortical cells are responsible for orientation discrimination, if not for the oblique effect itelf. Some studies have found differences that do seem capable of explaining an oblique effect. Pettigrew, Nikara, and Bishop (1968) studied cells with receptive fields located within 5 degrees of the visual axis (area centralis). These experimenters classified cells into several different categories. One kind of unit, a "simple cell" directionally sensitive to movement and giving just a single excitatory response, is of particular interest here. Among these cells, a distinct preference for vertical and horizontal orientations was found. This offers at least initial support for a neurophysiological explanation of the oblique effect. In the rabbit (Levick, 1965, 1967) ganglion cells have been found that respond to a horizontal or a vertical border, but none responding to obliques. Since rabbits can discriminate between oblique stimuli, these results suggest that these animals must use some other mechanism in making oblique discriminations.

### *Histological Evidence*

If neurons responsible for orientation analysis favor a particular orientation, it would be expected that some anatomical organization must provide a basis for this preference. Several studies suggest that the dendritic fields of sensory neurons provide such a code. In the optic lobe of the octopus, for example, cell dendrites have elongated oval fields with axes predominantly found in the horizontal plane, and to a lesser extent in the vertical (Young, 1960, 1962). Few fields have been found in oblique orientations.

Colonnier (1964) examined dendritic field orientation in cat, rat, and monkey cortex. He found that the number of dendritic segments was slightly greater in the horizontally or vertically oriented fields. The length of these segments was much greater than for oblique fields. Vertically oriented dendritic fields were especially predominant. Such an organization could account for differences in perception along different meridians.

> The orientation of elongated dendritic trees permits a differential sensitivity of the receptor cells to groups of differently oriented axons. . . . The preferential orientation of dendritic trees along one axis would ensure that a specific input coming in the same direction would be better focused upon the cells of that area [Colonnier, 1964, p. 341].

Colonnier suggested these preferences could be the morphological basis of an oblique effect. Although enough obliquely oriented fibers exist to make oblique discriminations possible, they should be less accurate than other discriminations.

### *Man*

Before going directly to a discussion of neurophysiological mechanisms in man, it should be pointed out that many investigators have tried to relate orientation preferences in man to optical factors. The earliest investigators attributed the results to astigmatism. Later studies, however, took steps to eliminate astigmatic contributions. The persistence of the preferences and consistency of astigmatism along two meridians argued against this explanation. Still, Beck (1965) showed that accommodation could lead to astigmatism for the near distances even where no astigmatism could be detected in the relaxed eye. Howard and Templeton (1966) have also criticized earlier studies for their inability to definitely eliminate astigmatism as a cause of the oblique effect.

Meanwhile, other optical theories had also been advanced. Leibowitz (1953) studied the role of dioptric factors by determining threshold angle resolution for test gratings in different orientations. Horizontal and vertical thresholds were found to be not different from each other, but markedly superior (by 7%) to all other orientations. Leibowitz reasoned that if dioptric aberrations were responsible for such results, the oblique effect should be enhanced under conditions where dioptric aberration was made more effective. However, the differences persisted even under these conditions. Leibowitz (1953) felt the data required a nondioptric explanation.

A later study on localization accuracy (Leibowitz, Myers, & Grant, 1955) supported this conclusion. The authors found that the detection of a patch of light was influenced by luminance and duration, but localization accuracy was not. Thus, the factors limiting localization performance were not governed by the photochemical events in the retina. Leibowitz et al. suggested this "points to the locus of contributing factors in events occurring after the activation of the end organs [p. 78]."

Weymouth (1959), however, persisted in explaining the oblique effect in optical terms. The characteristic effect of a point of light to form an image with horizontal and vertical rays (as in viewing a star) was argued to accentuate or increase the contrast of vertical and horizontal stimuli, while blurring an oblique.

Despite the appeal of these theories, recent studies have shown all theories based on the image-forming properties of the eye (astigmatism, diffraction characteristics, etc.) to be untenable. Campbell, Kulikowski, and Levinson (1966) effectively bypassed dioptric effects by projecting images directly onto the retina by means of laser-generated sinusoidal interference fringes. These are high-contrast gratings that are not degraded by ordinary optical effects. Resolving power was found to be worst for oblique patterns and best for vertical and horizontal gratings. Mitchell, Freeman, and Westheimer (1967) confirmed these results using the same technique. Both research groups concluded that orientation differences cannot be attributed to optical factors, and that retinal or higher mechanisms must be involved. Shlaer (1937) and Hartridge (1947) have argued for retinal mechanisms. They have suggested that density of retinal cells might be greatest along the horizontal and vertical meridians, and that this could lead to a corresponding enhancement in the perception of similarly aligned stimuli. This explanation, however, is not in agreement with histological evidence (see Polyak, 1941).

Still other investigators have ascribed the oblique effect to differential eye movements. Brown (1949) suggested that an asymmetrical distribution in the pattern of eye movements (resulting from preferential vibrations in the vertical and horizontal planes) may cause blurring along certain directions due to retinal image motion. Higgins and Stultz (1950) tested this hypothesis by presenting striped test objects to subjects at durations too short for eye movements to occur. In spite of the absence of eye movements, acuity for vertical and horizontal patterns was 20% better than for oblique objects. Nachmias (1960) recorded eye movements during threshold measurements and determined that they made no contribution at all to the observed differences.

In summary, investigators have tried to account for the oblique effect in terms of eye movements, optical disorders, and various dioptric characteristics. The composition of the retinal mosaic has also been proposed as a cause. Experimental studies show all of these explanations to be unacceptable, and therefore point to a postretinal origin. The evidence for this position is presented in the following sections.

### *Psychophysical Evidence*

One promising psychophysical approach to the study of neurophysiological mechanisms is visual masking (see Weisstein, 1969). Sekular (1965), for example, studied adaptation using a backward masking paradigm. Detection thresholds were studied as a function of the masking field orientation. It was found that detection of the test stimulus decreased as the masking stimulus aproached the orientation of the test field. Beyond 45 degrees the masking effect appeared to be independent of orientation. Similar findings are reported by Houlihan and Sekular (1968) and Parlee (1969). Blakemore, Nachmias, and Sutton (1970) have reported orientation-specific masking for gratings of similar spatial frequencies.

Since effective masking suggests an inhibitory or interfering effect by certain units on the activity of others, these results imply the presence of separate sets of analyzers for different orientations. This is supported by other orientation-specific masking phenomena such as in direction-sensitive analyzers (Sekular, Rubin, & Cushman, 1968) and orientation-specific color aftereffects (Fidell, 1970; Harris & Gibson, 1968; Hepler, 1968; McCollough, 1965).

The fact that a number of similar psychophysical investigations have discovered preferences for horizontal and vertical orientations suggests these systems are responsible for the oblique effect as well. Campbell and Kulikowski (1966) found that when a 45-degree test grating was used in place of a vertical field, angular selectivity as measured by the masking effect was reduced by 25%. Gilinsky (1968) and Gilinsky and Doherty (1969) also reported a different masking effect for obliques, while Bouma and Andriessen (1970) reported similar differences between obliques and horizontal or vertical lines for an angular interaction effect.

### *Neurophysiological Evidence*

The psychophysical studies support a neurophysiological theory for orientation perception and for the oblique effect in humans. More direct evidence, however, is available from some neurophysiological studies. Blakemore and Campbell (1969a, 1969b) reported that after 30 seconds of viewing an adapting grating of high contrast, a fivefold rise in threshold is obtained for a similarly oriented test object. This is accompanied by a suppression of the visual evoked potential, which only slowly returns to a detectable level. Since the visual evoked potential measures averaged activity of cortical cells, the source of this effect seems to be in the cortex.

Other studies have also found differences in the evoked potential to different orientations. Campbell and Maffei (1970) found that the amplitude of the evoked potential was .3 log units less for viewing an oblique test pattern than for vertical or horizontal patterns. This correlated with psychophysically derived thresholds for spatial frequency and contrast. Maffei and Campbell (1970) simultaneously recorded the evoked potential and the electroretinogram, thereby obtaining measures of electrical activity at both retinal and cortical levels. They found no differences in wave form or amplitude in the electroretinogram among vertical, horizontal, and oblique patterns. However, the evoked potential was decreased for the oblique patterns. Maffei and Campbell concluded:

> that there is an electrophysiological correlate to the psychophysical observation that the visual resolving power in oblique orientations is less than in the vertical and horizontal, and that the mechanism of this phenomenon arises between the site of origin of the electroretinogram and the evoked potential from the visual cortex [p. 387].

The most direct evidence for cortical analysis of orientation in man comes from Marg, Adams, and Rutkin (1968). These investigators were able to record activity from five cortical units in two conscious patients with intractable seizures. The subjects fixated a point on a screen while the receptive fields were plotted with moving bar-shaped or disc stimuli. Two of the receptive fields were circular and the others were rectangular. Among the rectangular fields, two were horizontal and the third was nearly vertical in orientation (95 degrees). Five units is hardly a sufficient sample on which to base conclusions, but the finding of only horizontal and vertical preferences certainly does not detract from other psychophysical and neurophysiological evidence that the oblique effect has a central origin in man.

## DISCUSSION

A review of the literature has shown that perception and discrimination vary with stimulus orientation. The most characteristic finding in these studies is a preference for vertical or horizontal orientations over obliques. This preference manifests itself in the resolution of targets, estimation of stimulus position, learning and discrimination of objects, and a wide assortment of other perceptual phenomena. These preferences are found throughout the animal kingdom, from man to octopus.

Neurophysiological research has firmly established the presence of cells in the cortex of cat, monkey, and man that are selectively sensitive to orientation. Ganglion cells in rabbit, pigeon, and goldfish have been found that perform similar functions. Histological evidence from octopus, rat, cat, and monkey suggest an anatomical substrate for orientation preferences.

It has become increasingly more popular to account for orientation preferences as reported in the behavioral literature in terms of these neurophysiological discoveries. Typically, it is assumed that preferences for horizontal and vertical orientations imply a disproportionate number of cells optimally sensitive to stimuli in these preferred positions. The distribution of preferred orientation does seem to be disproportionate in optic fibers of octopus, rabbit ganglion cells, and at least some subcategories of units in cat visual cortex. An anatomical basis for a bias in the orientation systems of rat, cat, and monkey has been found histologically. In man, psychophysical, evoked potential, and microelectrode evidence converge to suggest biases in the human visual system as well. In all cases the biases or disproportional distributions of analyzers have always favored the horizontal and vertical orientations at the expense of obliques.

These findings suggest that the oblique effect is a function of the number of cells responsible for analyzing stimuli in vertical, horizontal, and oblique positions. However, this is not a necessary assumption. The oblique effect can be accounted for in terms of differential sensitivity as well as distribution. So far, this theory has not been tested neurophysiologically. Two lines of inquiry seem useful in evaluating these positions. First, selectivity in single cells should be studied not only in terms of response maxima but in regard to the response distribution to a range of orientations. This would reveal whether cells preferring horizontal or vertical orientations are more "finely tuned" than obliques. Andrews (1967a, 1967b) has suggested such a mechanism. Second, a more thorough investigation should be made for any possible biases in the distribution of cell orientation preferences. Hubel and Wiesel (1968), for example, have admitted that their investigations were not extensive enough to reveal small predominances in any one orientation. It would also seem advisable to conduct a more extensive investigation of receptive fields in the peripheral retina. Only a few of these areas have been mapped to any degree. Distributions in the periphery might prove to be different from those encountered in more central regions.

Evidence continues to accumulate, and more specific data on these speculations should be forthcoming. But the data certainly suggest now that preferences for horizontal and vertical orientations, as reported in the behavioral literature, correspond to similar biases in neurological orientation-analyzing systems. The site of the oblique effect would seem to be well established in the higher visual pathways of man and animal. However, it would be remiss to imply that all aspects of this problem can be most successfully described in sensory or neurophysiological terms. For example, memorial or attentional components may exist, and these also deserve further study. In fact, if the suggested neurophysiological substrates of the oblique effect are real, the presence of attentional or memorial deficits would be expected. On the other hand, these factors could be important independent (in the absence) of neurophysiological biases.

## REFERENCES

ANDREWS, D. P. Perception of contours in the central fovea. *Nature,* 1965, **205**, 1218–1220.

ANDREWS, D. P. Perception of contour orientation in the central fovea. I. Short lines. *Vision Research,* 1967, **7**, 975–997. (a)

ANDREWS, D. P. Perception of contour orientation in the central fovea. II. Spatial integration. *Vision Research,* 1967, **7**, 999–1013. (b)

ATTNEAVE, F., & OLSON, R. K. Discriminability of stimuli varying in physical and retinal orientation. *Journal of Experimental Psychology,* 1967, **74**, 149–157.

ATTNEAVE, F., & OLSON, R. K. What variables produce similarity grouping. *American Journal of Psychology,* 1970, **83**(1), 1–21.

ATTNEAVE, F., & REID, K. W. Voluntary control of frame of reference and slope equivalence under head rotation. *Journal of Experimental Psychology,* 1968, **78**, 153–159.

BEALE, I. L., & CORBALLIS, M. C. Beak shift: An explanation for interocular mirror-image reversal in pigeons. *Nature,* 1968, **220**, 82–83.

BECK, J. Accommodative astigmatism and pattern acuity. *Journal of the Optical Society of America,* 1965, **55**, 1139–1142.

BECK, J. Effect of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics,* 1966, **1**, 300–302. (a)

BECK, J. Perceptual grouping produced by changes in orientation and shape. *Science,* 1966, **154**, 538–540. (b)

BECK, J. Perceptual grouping produced by line figures. *Perception & Psychophysics,* 1967, **2**, 491–495.

BLAKEMORE, C., & CAMPBELL, F. W. On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology,* 1969, **203**, 237–260. (a)

BLAKEMORE, C., & CAMPBELL, F. W. Adaptation to spatial stimuli. *Journal of Physiology,* 1969, **200**, 11–13P. (b)

Blakemore, C., & Cooper, G. F. Development of the brain depends on the visual environment. *Nature,* 1970, **228**, 477–478.

Blakemore, C., Nachmias, J., & Sutton, P. The perceived spatial frequency shift: Evidence for frequency selective neurons in the human brain. *Journal of Physiology,* 1970, **210**, 727–750.

Bouma, H., & Andriessen, J. J. Perceived orientation of isolated line segments. *Vision Research,* 1968, **8**, 493–507.

Bouma, H., & Andriessen, J. J. Induced changes in the perceived orientation of line segments. *Vision Research,* 1970, **10**, 333–349.

Brown, W. R. Eye movements and the variation of visual acuity with test object orientation. *Journal of the Optical Society of America,* 1949, **39**, 1057A.

Bryant, P. E. Perception and memory of the orientation of visually presented lines by children. *Nature,* 1969, **224**, 1331–1332.

Burns, B. D., Mandl, G., Pritchard, R., & Webb, C. The perception of briefly exposed point-sources of light. *Quarterly Journal of Experimental Psychology,* 1969, **21**, 299–311.

Campbell, F. W., & Kulikowski, J. J. Orientation selectivity of the human visual system. *Journal of Physiology,* 1966, **187**, 437–445.

Campbell, F. W., Kulikowski, J. J., & Levinson, J. The effect of orientation on the visual resolution of gratings. *Journal of Physiology,* 1966, **187**, 427–436.

Campbell, F. W., & Maffei, L. Electrophysiological evidence for the existence of orientation and size detectors in the human visual system. *Journal of Physiology,* 1970, **207**, 635–652.

Colonnier, M. The tangential organization of the visual cortex. *Journal of Anatomy* (London), 1964, **98**, 327–344.

Craig, E. A., & Lichtenstein, M. Visibility-invisibility cycles as a function of stimulus-orientation. *American Journal of Psychology,* 1953, **66**, 554–563.

Dodwell, P. C. Anomalous transfer effects after pattern discrimination training in rats and squirrels. *Journal of Comparative and Physiological Psychology,* 1970, **71**, 42–51.

Emsley, H. H. Irregular astigmatism of the eye: Effect of correcting lenses. *Transactions of the Optical Society,* 1925, **27**, 28–41.

Fidell, L. S. Orientation specificity in chromatic adaptation of human "edge-detectors." *Perception & Psychophysics,* 1970, **8**, 235–237.

Foley, P. J. Stimulus orientation and retinal summation. *Journal of the Optical Society of America,* 1962, **52**, 474–475.

Ganz, L., & Fitch, M. The effect of visual deprivation on perceptual behavior. *Experimental Neurology,* 1968, **22**, 638–660.

Ganz, L., Fitch, M., & Satterberg, J. A. The selective effect of visual deprivation on receptive field shape determined neurophysiologically. *Experimental Neurology,* 1968, **22**, 614–637.

Gibson, E. J., Gibson, J. J., Pick, A. D., & Osser, H. A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 897–906.

Gilinsky, A. Orientation-specific effects of patterns of adapting light on visual acuity. *Journal of the Optical Society of America,* 1968, **58**, 13–18.

Gilinsky, A. S., & Doherty, R. S. Interocular transfer of orientation effects. *Science,* 1969, **164**, 454–455.

Green, R. T., & Hoyle, E. M. The influence of spatial orientation on the Poggendorf illusion. *Acta Psychologica,* 1964, **22**, 348–366.

Hamblin, J. R., & Winser, T. H. On the resolution of gratings by the astigmatic eye. *Transactions of the Optical Society,* 1927, **29**, 28–42.

Harris, C. S., & Gibson, A. R. Is orientation-specific color adaptation in human vision due to edge detectors, afterimages, or dipoles? *Science,* 1968, **162**, 1506–1507.

Hartridge, F. R. The visual perception of fine detail. *Philosophical Transactions of the Optical Society* (London), 1947, **B232**, 614.

Hepler, N. Color: A motion-contingent aftereffect. *Science,* 1968, **162**, 376–377.

Higgins, G. C., & Stultz, K. Visual acuity as measured with various orientations of a parallel-line test object. *Journal of the Optical Society of America,* 1948, **38**, 756–758.

Higgins, G. C., & Stultz, K. Variation of visual acuity with various test-object orientations and viewing conditions. *Journal of the Optical Society of America,* 1950, **40**, 135–137.

Hirsch, H. V., & Spinelli, D. N. Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science,* 1970, **168**, 869–871.

Horn, G., & Hill, R. M. Modifications of receptive fields of cells in the visual cortex occurring spontaneously and associated with bodily tilt. *Nature,* 1969, **221**, 186–188.

Houlihan, K., & Sekular, R. W. Contour interactions in visual masking. *Journal of Experimental Psychology,* 1968, **77**, 281–285.

Howard, I. P., & Templeton, W. B. *Human spatial orientation.* London: Wiley, 1966.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology,* 1962, **160**, 106–154.

Hubel, D. H., & Wiesel, T. N. Shape and arrangement of columns on cat's striate cortex. *Journal of Physiology,* 1963, **165**, 559–568.

Hubel, D. H., & Wiesel, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 & 19) of the cat. *Journal of Neurophysiology,* 1965, **28**, 229–289 .

Hubel, D. H., & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology,* 1968, **195**, 215–243.

Huttenlocher, J. Discrimination of figure orientation. *Journal of Comparative and Physiological Psychology,* 1967, **63**, 359–361.

Jacobson, M., & Gaze, R. M. Types of visual response from single units in the optic tectum and optic nerve of the goldfish. *Quarterly Journal of Experimental Physiology,* 1964, **49**, 199–209.

Jastrow, J. On the judgment of angles and positions of lines. *American Journal of Psychology,* 1893, **5**, 214–248.

Judd, C. H., & Courten, H. C. The Zöllner illusion. *Psychological Monographs,* 1905, **7**, 112–141.

Koronakos, C., & Arnold, W. J. The formation of learning sets in rats. *Journal of Comparative and Physiological Psychology,* 1957, **50**, 11–14.

Lashley, K. S. The mechanisms of vision: XV. Preliminary studies of the rat's capacity for detail vision. *Journal of General Psychology,* 1938, **18**, 123–193.

Leibowitz, H. Some observations and theory on the variation of visual acuity with the orientation of the test object. *Journal of the Optical Society of America,* 1953, **43**, 902–905.

Leibowitz, H. Some factors influencing the variability of vernier adjustments. *American Journal of Psychology,* 1955, **68**, 266–273.

Leibowitz, H. W., Myers, N. A., & Grant, D. A. Radial localization of a single stimulus as a function of luminance and duration of exposure. *Journal of the Optical Society of America,* 1955, **45**, 76–78.

Leibowitz, H., & Toffey, S. The effect of rotation and tilt on the magnitude of the Poggendorf illusion. *Vision Research,* 1966, **6**, 101–103.

Levick, W. R. Receptive fields of rabbit retinal ganglion cells. *American Journal of Optometry,* 1965, **42**, 337–343.

Levick, W. R. Receptive fields and trigger features of ganglion cells in the visual streak of the rabbit's retina. *Journal of Physiology,* 1967, **188**, 285–307.

Luria, S. M. The effect of body-position on meridional variations on scotopic acuity. *American Journal of Psychology,* 1963, **76**, 598–606.

Mackintosh, J., & Sutherland, N. S. Visual discrimination by the goldfish: The orientation of rectangles. *Animal Behavior,* 1963, **11**, 135–141.

Maffei, L., & Campbell, F. W. Neurophysiological localization of the vertical and horizontal visual coordinates in man. *Science,* 1970, **167**, 386–387.

Marg, E., Adams, J. E., & Rutkin, B. Receptive fields of cells in the human visual cortex. *Experientia,* 1968, **24**, 348–350.

Maturana, H. R., & Frenk, S. Directional movement and horizontal edge detectors in the pigeon retina. *Science,* 1963, **142**, 977–979.

McCollough, C. Color adaptation of edge-detectors in the human visual system. *Science,* 1965, **149**, 1115–1116.

McFarland, J. H. "Parts" of perceived visual forms: New evidence. *Perception & Psychophysics,* 1968, **3**, 118–120.

Mello, N. Inter-hemispheric reversal of mirror-image oblique lines after monocular training in pigeons. *Science,* 1965, **148**, 252–254.

Mello, N. Interocular generalization: A study of mirror-image reversal following monocular discrimination training in the pigeon. *Journal of the Experimental Analysis of Behavior,* 1966, **9**, 11–16. (a)

Mello, N. Concerning inter-hemispheric transfer of mirror-image patterns on pigeon. *Physiology and Behavior,* 1966, **1**, 293–300. (b)

Mello, N. Inter-hemispheric comparison of visual stimuli in the pigeon. *Nature,* 1967, **214**, 144–145.

Mitchell, D. E., Freeman, R. D., & Westheimer, G. Effect of orientation on the modulation sensitivity for interference fringes on the retina. *Journal of the Optical Society of America,* 1967, **57**, 246–249.

Nachmias, J. Meridional variations in visual acuity and eye movements during fixation. *Journal of the Optical Society of America,* 1960, **50**, 569–571.

Nissen, H. W., & McCulloch, T. L. Equated and non-equated stimulus situations in discrimination learning by chimpanzees. *Journal of Comparative Psychology,* 1937, **23**, 165–189.

Ogilvie, J. C., & Taylor, M. M. Effect of orientation on the visibility of fine wires. *Journal of the Optical Society of America,* 1958, **48**, 628–629.

Ogilvie, J. C., & Taylor, M. M. Effect of length on the visibility of a fine line. *Journal of the Optical Society of America,* 1959, **49**, 898–900.

Onley, J. W., & Volkman, J. The visual perception of perpendicularity. *American Journal of Psychology,* 1958, **71**, 504–516.

Over, R. Detection and recognition measures of shape discrimination. *Nature,* 1967, **214**, 1272.

Over, R., & Over, J. Detection and recognition of mirror-image obliques by young children. *Journal of Comparative and Physiological Psychology,* 1967, **64**, 467–470.

Parlee, M. B. Visual background masking of a single line by a single line. *Vision Research,* 1969, **9**, 199–205.

Parriss, J. R. A technique for testing cat's discrimination of differently oriented rectangles. *Nature,* 1964, **202**, 771–773.

Pettigrew, J. D., Nikara, T., & Bishop, P. O. Responses to moving slits by single units in cat striate cortex. *Experimental Brain Research,* 1968, **6**, 373–390.

Polyak, S. L. *The retina.* Chicago: University of Chicago Press, 1941.

Rochlin, A. M. The effect of tilt on the visual perception of parallelness. *American Journal of Psychology,* 1955, **68**, 223–236.

Rudel, R. G., & Teuber, H.-L. Discrimination of direction of line in children. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 892–898.

Sekular, R. W. Spatial and temporal determinants of visual backward masking. *Journal of Experimental Psychology,* 1965, **70**, 401–406.

Sekular, R. W., & Rosenblith, J. F. Discrimination of direction of line and the effect of stimulus alignment. *Psychonomic Science,* 1964, **1**, 143–144.

Sekular, R. W., Rubin, E. L., & Cushman, W. H. Selectivities of human visual mechanisms for direction of movement and contour orientation.

*Journal of the Optical Society of America*, 1968, **58**, 1146–1150.

Shlaer, S. The relation between visual acuity and illumination. *Journal of General Physiology*, 1937, **21**, 165–188.

Smith, S. I. Angular estimation. *Journal of Applied Psychology*, 1962, **46**, 240–246.

Sulzer, R. L., & Zener, K. A quantitative analysis of relations between stimulus determinants and sensitivity of the visual perception of parallelness. *American Journal of Psychology*, 1953, **8**, 444.

Sutherland, N. S. Visual discrimination of orientation by octopus. *British Journal of Psychology*, 1957, **48**, 55–71.

Sutherland, N. S. Visual discrimination of the orientation of rectangles by octopus vulgaris lamark. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 452–458.

Sutherland, N. S. Cat's ability to discriminate oblique rectangles. *Science*, 1963, **139**, 209–210. (a)

Sutherland, N. S. Shape discrimination and receptive fields. *Nature*, 1963, **197**, 118–122. (b)

Taylor, M. M. Visual discrimination and orientation. *Journal of the Optical Society of America*, 1963, **53**, 763–765.

Van Hof, M. W. Discrimination between striated patterns of different orientation in the rabbit. *Vision Research*, 1965, **6**, 89–94.

Van Hof, M. W. The mechanism of orientation discrimination in the rabbit. *Pflügers Archiv*, 1969, **311**, 195–196.

Van Hof, M. W. Interocular transfer in the rabbit. *Experimental Neurology*, 1970, **26**, 103–108. (a)

Van Hof, M. W. Mechanisms of orientation discrimination in the rabbit. *Experimental Neurology*, 1970, **28**, 494–500. (b)

Van Hof, M. W., & Wiersma, C. A. The angular threshold of discrimination for striated patterns of different orientation in the rabbit. *Vision Research*, 1967, **7**, 265–270.

Wallace, G. K. Measurements of the Zöllner illusion. *Acta Psychologica*, 1964, **22**, 407–412.

Warren, J. M. Discrimination of mirror-images by cats. *Journal of Comparative and Physiological Psychology*, 1969, **69**, 9–11.

Weisstein, N. What the frog's eye tells the human brain: Single cell analyzers in the human visual system. *Psychological Bulletin*, 1969, **72**, 157–176.

Wells, M. J. Proprioception and visual discrimination of orientation in octopus. *Journal of Experimental Biology*, 1960, **37**, 489–499.

Weymouth, F. W. Stimulus orientation and threshold: An optical analysis. *American Journal of Opthalmology*, 1959, **48**, 6–10.

Wiesel, T. N., & Hubel, D. H. Extent of recovery from the effects of visual deprivation in kittens. *Journal of Neurophysiology*, 1965, **28**, 1060–1072.

Young, J. Z. The visual system of octopus. *Nature*, 1960, **186**, 836–839.

Young, J. Z. Memory mechanisms of the brain. *Journal of Mental Science*, 1962, **108**, 119–133.

Zeigler, P., & Schmerler, S. Visual discrimination of orientation by pigeons. *Animal Behavior*, 1965, **13**, 475–477.

(Received February 19, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 4, 279–291

# RECONSIDERATION OF A MEASURE OF POWER MOTIVATION[1]

JOSEPH VEROFF AND JOANNE B. VEROFF[2]

*University of Michigan*

A review of research using a thematic apperceptive measure of power motivation suggests that it largely reflects a person's fear of weakness. Evidence for this assertion comes from demographic distributions of the measure, and from its associations with feelings of adjustment in specific and general life roles. Such a fear of weakness seems to lead to avoidant power behaviors in some instances, some cost in adjustment in other instances, and a positive adjustment in still other instances. A context for thinking about power motivation as an orientation to a negative goal is provided.

The problem of man's strivings for power has held a perpetual and pervasive fascination for students of the human race. As the relationships between individuals, between man and society, and between societies grow increasingly complex, the problem becomes increasingly important to understand. One systematic, empirical approach to this area has been the use of an apperceptive measure of power motivation (Atkinson, 1958). A number of researchers have used this measure in social investigations over the past 15 years. The results of these investigations point to some themes about the kind of power motivation this instrument measures, both with respect to its significance in adult striving and to the way in which it is developed in human interaction. It seems useful at this time to summarize these themes and to attempt a theoretical integration of them, with the hope that any modification in a theory of social motivation that includes a discussion of power motivation might be responsive to available strivings systematic research using this measure.

Because of the many nuances of meaning that are attributed to the word "power," it is important to begin by specifying what the apperceptive measure of power motivation was intended to measure. Some theorists and researchers concerned with the problem of power have defined it as a sense of effectiveness within the individual. Minton (1967) has called this kind of power "latent power" and has provided an extensive review of relevant studies. Others have explored the realm of power by looking at these individuals who manifestly possess power (see Schopler, 1965). We, however, have been primarily interested in the *desire* for power in terms of influence and control. As originally constructed (Veroff, 1958), the measure was designed to estimate power motivation through fantasy which was coded for "evidence of concern with attaining or maintaining control of the means of influence." Imagery in stories was coded for evidence of power motivation if some affect was expressed over gaining, maintaining, or losing influence; if status differences were accentuated by role relationships between characters; and if interpersonal influence was clearly indicated to have occurred. In its initial formulation, the measure distinguished college students invested in running for student office from those who were not; argumentative students from those who were not argumentative in the classroom; students low in social values on the Allport-Vernon scale of values from those who scored higher; students with older siblings from those without any older siblings; students who ranked leadership as an important aspect of job satisfaction relatively high from those who ranked it relatively low. From these comparisons the significance of power motivation seemed to reside in general concern with being an influencer without any further specification of the basis of this concern or the conditions that would likely engage this concern. Since this beginning work, the implicative meaning of the measure has expanded in the course of subsequent correlational studies using the measure. Many different pieces of evidence support the view that the power incentive defined as "control of influence of others" should not be generally conceived as a *positive* goal in which there is active joy from having impact per se but primarily as a *negative* goal in which there is, as Adler suggested, an avoidance of feelings of powerlessness as men strive for perfection. Since this is the most important conclusion of this paper, clarification of the terms, positive and negative goals, is in order.

---

[1] Writing this paper was facilitated by a United States Public Health Service Special Fellowship to Joseph Veroff during the year 1967–1968 at the Institute of Human Development, Berkeley, California. The director of the Institute at that time, M. Brewster Smith, was very helpful in going over a previous draft of this paper, as was Daniel Katz.

[2] Requests for reprints should be sent to Joseph Veroff, Department of Psychology, University of Michigan, Ann Arbor, Michigan 48104.

We have generally followed an expectancy-value position on power motivation. Derived from Lewin, Dembo, Festinger, and Sears (1944), Tolman (1955), and Atkinson (1964), our model makes no assumption about a closed system of energy—as can be found in Adler's (1956) thinking about power motivation as compensation for developmental interference with man's striving for perfection—nor does our model make any assumption about a person's intrinsic tendency to assert his own force—as can be found in White (1959), in Smith's (1968) restatement of White, and in deCharms (1968). Rather, the model suggests that the person's phenomenal field is made up of negative and positive goals that are the *anticipations about the consequences of power-instrumental activities that depend on the frequency of punishing and rewarding experiences with these activities*. Thus the goal either positive or negative may be stronger in one situation or another, depending on past associations to that situation. Furthermore, the model suggests that these goals are mediated by a person's motives, his stable affective orientation to power goals.

It is our contention that the power motive measure that is discussed in this paper largely reflects a generalized concern about negative power goals, a concern we have called fear of weakness. There is nothing, however, about the data we discuss that suggests that the expectancy-value framework best accounts for the results. Different theoretical positions seem to be useful in understanding different sets of results. We point out the special usefulness of a different theoretical position as it occurs.

## SOME METHODOLOGICAL CAUTIONS

Before we look at this evidence we would like to enter a cautionary note about the particular assessment technique used in most of the research reported in this paper. The research, by and large, employs the thematic apperceptive assessment of power motivation or a derivation of such assessment. The apperceptive assessment of power motivation depends on coding stories that are told in response to pictures for imagery reflecting concern over the means of influence. The coding reliability is reasonable but not as good as the reliability with achievement motivation (Feld & Smith, 1958).

Furthermore, apperceptive measurement in general has not held up well in studies of stability of the measurement over time; there have been only low positive correlations in test–retest studies (see Feld, 1967). However, this measure being a test of imagination suggests to people that they should not repeat the theme of the story they have previously told but select a theme representing a response lower in their hierarchy of usual responses.

Another reservation about the measure that ought to be raised is with regard to its unidimensionality. Undoubtedly, there ought to be not one measure called power, but several. Weinstein (1969) has recently questioned the unidimensionality in the concept of achievement motivation as reflected in fantasy, since there is very little convergent validity found in a factor-analytic treatment of the measure along with other measures. The same question undoubtedly can be raised about the fantasy measure of power. Uleman's (1972) and Winter's (1971) work on different ways of measuring the power motive also raise the same problems of unidimensionality.

In spite of these shortcomings, however, we think there is an impressive array of research stemming from the use of apperceptive devices in relation to more specified motivational measures. The measure has had unusually high heuristic value. We hope that with this paper we can encourage use of the measure not only for theoretical amplification but also to elicit methodological refinements that will improve the admittedly crude procedure.

Let us turn now to research findings, which can be grouped under three main sources: (*a*) demographic distributions of the measure of power motivation in the American social structure, (*b*) reactions of men and women with high power motivations to being in certain life roles, (*c*) relationships to general social adjustment. We consider each of these sources separately.

## DEMOGRAPHIC DISTRIBUTIONS OF POWER MOTIVATION

Veroff, Atkinson, Feld, and Gurin (1960) discussed the methodological and interpretative problems encountered when a thematic apperceptive measure was administered in a nationwide sample survey, the major goal of which was to describe the feelings of subjective adjustment in the American population. A sample of 1,619 men and women, representative of the adult population of the United States in 1957, told stories in response to pictures specifically chosen to elicit not only power imagery but also achievement and affiliation imagery as well. These protocols were scored for power motivation according to a system worked out by Veroff (1957) that had been previously validated by experimental and correlational studies.

From the distribution of the power motive scores in various subsamples of this adult population, we can get some insights into the meaning of the power motive measure. Generally speaking, the distinction and differences between mean power motive scores among subgroups of the American population were not large; however, there were enough consistent trends to indicate a dominant interpretation of the measure; the power motive measure should be interpreted as a *concern about weakness*.

Two of these results are presented here. Table 1 depicts the percentages of high power motive scores[3] among men and women with different educational attainment. We find that

TABLE 1

PERCENTAGE OF HIGH POWER MOTIVE SCORES AMONG MEN AND WOMEN WITH DIFFERENT EDUCATIONAL ATTAINMENT

| Highest educational level attained | Men | | Women | |
|---|---|---|---|---|
| | % High power motive | $N$ | % High power motive | $N$ |
| Grade school | 55 | 176 | 42 | 222 |
| High school | 47 | 271 | 51 | 422 |
| College | 49 | 146 | 54 | 127 |

men who have gone only as far as grade school have higher power motive scores than those who have higher education. We find the reverse true for women. The explanation of the finding for men seems clear. Education is one prerequisite for successful career alignment, a means of attaining prestige in our society. A man who has not been able to acquire education is in a low position in our society and consequently experiences feelings of inferiority in contrast to other men. Another example can be cited: The highest power motive scores came from men who had incomes of under $2,000. As an even more dramatic example of this thesis, we found a very large difference between power motive scores of black and white men in our sample. For black men ($N = 31$) 61% scored above the median need power score, while this was true for only 48% of the white men ($N = 538$) in our sample. Weems and Wolowitz (1969) confirmed these findings. In comparing black and white schizophrenics, they found more black schizophrenics giving power themes depicted as "deficits, surpluses, or inequalities of personal resources related to masculine self-esteem [p. 189]." Wolowitz and Shorkey (1966) had previously found this imagery especially prevalent in male paranoia in general, which they view as a reflection of the paranoid's disastrous failure to meet the competitive demands of masculine assertion in the early family setting, a failure believed more commonly experienced by blacks in a system oppressive to black males.

What about women? The relationship between educational attainment and power motivation is reversed: It is not the women with low educational attainment but the women with high educational attainment who have high power motive scores. At first blush this seems to be an irreconcilable contradiction to the statement that power motivation reflects a deprivation in status. However, a more sophisticated view of the female's role in our society might suggest that in fact women with college education deprive themselves of usual feminine status in our society. They are influencing as *men*; they are not conforming to what women's role would traditionally prescribe for them. Thus, a woman with a college degree may suffer some feelings of inferiority with regard to her own status as a woman. This is further corroborated in other information from the nationwide survey; single women have the highest power motive scores of any other group of women. Of the 35 single women 60% were above the median power motive score. Single women in our society lack society's major prerequisites for the status of being a woman: the marital status. This group, thus, experiences strong feelings of inferiority and emerges with high power motivation scores. Another piece of evidence comes from comparing married women who do and do not have children. Women who do not have children have much higher power motive scores than women who do. Again we would take this result to demonstrate that a woman who does not bear society's stamp of being a totally adequate women is likely to have high power motivation.

---

[3] Here and in subsequent references this term signifies the percentage above total median.

Another way to interpret this result is in "relative deprivation" terms. Although college-educated women may have more status in overall societal perspective, by virtue of their education they become increasingly interested in playing a man's role in the society. Thus, they may evaluate their status not with respect to other women, but with respect to men. In such evaluation they may feel relatively more deprived of status than do women with less education who use only the status of women as their basis of evaluation.

In either case we see from these demographic analyses that men and women deprived of status have high power motivation. For men, to be deprived of socioeconomic status in our society seems to go along with high power motivation. For women, to be deprived of feminine or masculine status, to be considered inadequate in social terms either in the woman's world or in the man's world, seems to go along with high power motivation. These results then would lead us to believe that Adler's assumption that power motivation is precipitated by feelings of inferiority with regard to one's search for perfection and status is a legitimate position to take.

We have very little evidence from the demographic analysis that men and women in social roles of high status have high power motivation. Skolnick (1966), however, did report some interesting relationships between power motivation scores and leadership and status dimensions in high-school-age adolescents. She found high power motivation scores going along with high status and leadership roles. It perhaps should be noted that the power-motivation scores were obtained in this study from a coding of stories told to the Thematic Apperception Test (TAT)—an instrument heavily weighted with pictures whose content suggests negative interpersonal conflict. It is possible that the TAT pictures may arouse such intense concern about inadequacy and weakness that they elicit avoidant responses in people already so concerned, who then do not tell power stories at all. Under such circumstances those receiving high power motivation scores would be those more positively invested in power.

We prefer a different interpretation of Skolnick's results. We would suggest that in the high school adolescent's social perspective, power is not threatening, and therefore it is easy for him to translate a fear of weakness into overt power behaviors in social interaction. As background to the latter statement, let us examine the results of a study by Berlew (1961).[4] These researches find that high power motivation is associated with *interpersonal insensitivity under power-arousing conditions* in the laboratory, while moderate power motivation is associated with sensitivity. Berlew invoked the Yerkes-Dodson principle, which states that the most effective performance occurs at moderate motivation. Under very high motivating conditions the high power motive person will become ineffective. It is as if the consensually recognized clarity of the power goal in a situation does not give him enough leeway for attaining partial or subtle power goals *in his own terms*. And more important he cannot afford power failure, when everyone can recognize his weakness. When would high power motivation be effective? Under unthreatening power conditions, when the external conditions do not *clearly raise one's concern about power and status,* attaining and maintaining positive power *remain as possible goals*. Under these conditions, high power motivation is presumably not magnified to an intensity that interferes with social behavior. A person with high motivation can thus set his own standard of successful control, and thereby operate effectively and without anxiety.

---

[4] Berlew's study is important methodologically, for many of the theses we are about to propose. It, along with an unpublished study by Veroff, showing that high power-motivated subjects tended to influence more in joint decision making about food preference, and a study by Terhune (1968), dis-

Is this what is going on in Skolnick's adolescents? Is adolescence a time when power and status are just beginning to be crystallized as important social concerns? These goals are omnipresent in adolescence, but they are not major themes running through adolescents' *overt* behavior. They overtly concentrate on social belonging, struggles with role competence, and heterosexual love. Early adolescence thus may not evoke strong arousing conditions for power. Thus young adolescents with high power motivation might be very successful adolescents, those who are popular and effective leaders.

It is not until after some time that adolescents become clear about the fact that they are in fact struggling for autonomy and assertion, for some differentiation from their family and friends, in other words, for power and status. It is not until late adolescence that they have to make public vocational choices about their future in the social structure, although they were socialized into certain choices quite early in their school careers. A study by Johnson and Smith (1965) shows that among a large sample of adolescents there is a peculiarly high incidence of high power motivation in thematic protocols during *late* adolescence, rather than earlier. It is as if, as we are suggesting, the themes of power and status become crystallized in late adolescence and can be new socializing forces in the development of power motivation. If we are correct in our interpretation, we would expect that Skolnick's findings might begin to reverse during late adolescence, that power-motivated adolescents may begin to avoid power and become ineffective in social interaction.

Such a way of thinking about power motivation—that it induces successful overt power behavior only under relatively ambiguous conditions—has had some support in Andrews' (1967) recent work. Andrews looked at the motivations of successful executives in two different business firms, one having clear achievement and status standards, the other characterized by paternalism and a pecking order based on *both* dominance and dependence with no clear promotion standards. In the first firm, a man's overt power was laid on the line; in the second, he could get ahead both by being overtly power oriented and by being overtly not power oriented but rather ingratiating and affiliative. Success in the firm could result from participating in both activities, and thus an employee was never sure whether he was being evaluated for being a nice guy or for being influential. Andrews found that individual advancement of executives in each firm was related to the strength of both achievement and power motivation of the executives. In the "achievement" firm, in which status was clearly de-

---

cussed in another section, are the only ones we know in which power-motivated subjects were placed in an experimental situation and then observed. Only in these studies can the direction of causation between power motivation and another variable be confidently stated as power motivation influencing some phenomenon. Many of the studies reported here have the alternative causative possibility—that the given phenomenon affected the strength of power motivation. In general we assume that the power motivation measure reflects motivating properties, but we recognize that it also can reflect reactions to life situations as well. When both interpretations seem plausible to us, we present both.

fined through achievement, advancement in the firm was positively related to high achievement motivation and *negatively* to high power motivation. In the other firm these correlations were reversed. Thus, in firms in which advancement depended on confusing demands, with the person sometimes being asked to act ingratiatingly and sometimes asked to act overtly power oriented, the man with high power motivation apparently was able to succeed, while the man with high achievement motivation was not. Power motivation is thus ineffective or covert where status is clearly defined as a demand in one's work, but it becomes effective and overt when the demands for status are fuzzed up.

McKeachie (1961) reported results about highly power motivated students getting better grades in classes where student volunteering is prevalent than in classes where there is a clear formal teaching structure, undoubtedly where the mode of influence is clearly in focus all the time. Again we might interpret these results to mean that a power-motivated person operates effectively where the influence structure is loose or ambiguous and does not operate effectively where the structure is well defined.

In both Skolnick's and Andrews' studies we should be aware of the possibility that it was the fact of being in these role positions—a statusful adolescent, a successful member of a nonachievement organization, or an unsuccessful member of an achieving organization—that may have induced the higher measured power motivation. In all of these role positions, we might think of the person as being in an especially vulnerable public status that might *induce* an interest in power. We prefer the opposite interpretation, but we recognize that our derivations are ambiguously based.

## ROLE REACTIONS OF MEN AND WOMEN WITH HIGH POWER MOTIVATION

Similar conclusions about the measure of power motivation come from analyses of the reactions of people with high and low power motivation in different life roles. In *Marriage and Work in America* Veroff and Feld (1970) concluded that when the measure of power motivation is correlated with reaction to, and perceptions of, life roles, there is little evidence to indicate that men and women get pleasurable *positive* gratifications from power exertion in life roles that clearly demand power. From most of their findings, Veroff and Feld suggest that men and women with high power motivation are obtaining satisfaction from life roles, or perceiving certain aspects of life roles because of their attempts to *avoid feelings of weakness*. A number of illustrations of this point are in order.

Let us start with the work role. Lindman[5] had earlier found that high school boys who had high power motivation were more likely than high school boys with low power motivation to prefer those occupations that they perceive as either very high or very low in power status.

With Lindman's results as background we looked for confirmatory results in our analysis of the power motive in the job role in the national survey. There was a negative correlation between a man's power motivation and whether his job allowed the possibility for supervision of others ($\gamma = .41$, $p < .05$).[6] Men with high power motivation were less likely to be in jobs that had supervisory potential. Two particular occupations were characterized by men with very low power motive scores—salesmen and foremen. Only 38% of salesmen ($N = 29$) were above the median power motive score; only 24% of the foremen ($N = 33$). These are occupations that *do* offer considerable potential for demonstration of power and status and influence. Perhaps, as Lindman's results suggest, some men with high power motivation actually avoid these jobs.

Let us turn to another result about the occupational role that is discussed in *Marriage and Work in America*. Table 2 reports the relationship of satisfaction at work to power motivation at different levels of super-

---

[5] H. Lindman. A study of the motives and the vocational preferences of high school students. Unpublished honors thesis, University of Michigan, 1958.

[6] Gamma, a measure suggested by Goodman and Kruskal (1954), describes how consistently one can reproduce the order of one variable when the order of the other variable is known. It can be thought of as the percentage of increase in predicting the order of one variable knowing the other. In this case it would be 41%.

TABLE 2

Relationship of Satisfaction to Power Motivation at Different Levels of Supervision of the Respondent

| Amount of Supervision of Work × Power Motive Score | N | % Satisfied with job | % Dissatisfied with job |
|---|---|---|---|
| Much supervision | | | |
| High power motivation | 18 | 39 | 61 |
| Low power motivation | 24 | 79 | 21 |
| Little supervision | | | |
| High power motivation | 12 | 83 | 17 |
| Low power motivation | 17 | 71 | 29 |

vision of the respondent. Respondents were divided into two groups: those who said they were supervised a lot as opposed to those who said they were supervised just a little. Controlling for this level of supervision, we plotted how satisfied or dissatisfied men felt in their work if they had high as opposed to low power motivation. The results are reported for younger men only.[7] Table 2 shows that men with high power motivation in jobs with little supervision are the most satisfied. These results corroborate our view that men with high power motivation, being concerned about weakness, can have their feelings of inferiority actively aroused if they experience much supervision by others.

In the analysis of data about parental role we found considerable evidence for thinking that parenthood—a clearly power-relevant role—elicits feelings of inadequacy more often in parents when they have high power motivation than when they have low power motivation. This was especially true for women when their children were young, and especially true for men when their children were older. We asked parents whether they ever felt they were not as good parents as they would like to be. In mothers whose children were preschool age, we found significant relationships between power motive and admitting to some feelings of inadequacy as a parent while the opposite tended to be true for mothers whose children were older. Power-motivated women feel inadequate when their major functions of motherhood are open to public scrutiny: when they have to nurture and socialize *young* children. Power-motivated men feel inadequate when their children are adolescents and their major authority is open to public scrutiny. When they are launching a child into adulthood the consequences of their authority are very visible. Particularly interesting was the fact that as children grew older, both mothers and fathers with high power motivation tended to report feelings of affiliative inadequacies as parents (i.e., feeling they do not get to see their children as much as they would like to). As children grow older, there are more glaring potential disruptions to men's and women's feelings of power. A power-motivated parent, concerned about weakness, might tend to avoid contact with his child, especially as his child grows older. We found this pattern of results to be especially significant in our data among fathers with little social status—those with a grade school education. It is these men, who, as their children grow up, may find their parenthood most threatening to their own feelings of adequacy; indeed, it is this group of men who, when their children were adult, reported finding parenthood most problematic if they had high power motivation.

To go to another role—friendship—let us mention a recent study (Mann, 1968) that shows that college students with high power motivation early in their college career develop friendships with people they view as being inferior to them less frequently than do college students with low power motivation. It is difficult to know whether these people really surround themselves with people whom they view as not inferior to themselves or whether the high power-motivated individuals are inclined to see any friends in this way. The second explanation seems to be more reasonable, but it is tenuous.

Three studies of an entirely different sort take on new meanings if we assume that the power motivation measure is primarily assessing concern about inferiority in social roles. This time we are highlighting *general* societal roles rather than specific societal roles; we are focusing on men and women vis-à-vis their relationship to society in general. McClelland (1961) in his study of school textbooks of different countries from different times in history found that the emergence of dictatorships in countries was preceded by the use of textbooks whose content was high in power imagery and low in affiliation imagery. These results can be interpreted as follows: When there is a dominant concern over inferiority or weakness in a society, that society is more receptive to an emerging dictatorship. The textbooks themselves in focusing on power and influence, in effect, "teach" the children to have high power motivation or, as we have assumed, a strong concern about weakness. Such a concern may facilitate the development of a dictatorial style of government, because, if we are correct in our interpretations, power-motivated people often engage in power-avoidant behaviors and would not challenge the power of a dictator; they might be prone to escape from any freedom that was clearly labeled as a reflection of the person's own power.

---

[7] A similar analysis for older men did not yield any significant findings. We would interpret this to mean that there are possibilities of adaptation to both feelings of satisfaction and to the power motive that can occur for older men that are not open to younger men.

Rudin (1965) found a positive relationship between power motivation, again as measured in textbooks prevalent in 1925 in 17 different countries, and the subsequent incidences in 1950 of suicide, homicide, and alcoholism (as measured by the incidence of death due to cirrhosis of the liver). These results are consistent with the conceptualization of power motivation as fear of weakness. Certainly, alcoholism has been generally interpreted as a symptom of an overwhelming feeling of inability to cope with life's problems. And perhaps the ultimate level of concern over inadequacy of inferiority has been reached when a man perceives the taking of his own life or that of another as a solution to his problems in society.

Barrett and Franke (1970) have recently challenged some of Rudin's findings. They used the sample of textbooks as of 1950 and tried to replicate Rudin's findings in a larger sample of countries by predicting the incidence of homicide, suicide, and alcoholism as of 1965. The authors were discouraged by what they said was lack of replication, but a close examination of data suggests they might have been unduly negative. They *did* find a positive correlation between power motivation in children's readers in 1950 and homicide rates in 1965, repeating Rudin's earlier findings on this. There was no replication of the relationships between power motivation and suicide rates nor between power motivation and incidence of deaths due to cirrhosis of the liver, but in the latter case there was a switch from a positive correlation found in Rudin's study (.38) to a negative one in Barrett and Franke (– .45). This switch is remarkable and perhaps suggests that societies are moving away from using self-annihilation to homicide as a way to compensate for or to bolster feelings of inferiority. As rising expectations about the efficacy of power occur in the social world, in revolution and collective violence, then perhaps more inward expressions of inferiority are less likely to occur. These results are very provocative for thinking about power motivation in collective phenomena in the future.

These studies deal with the relationship between an aspect of a cultural product—school textbooks—and general societal problems rather than the behavior of individuals and the concern of individuals about power in their fantasy products, and, as such, they represent a very different realm of discourse. However, the findings are extremely provocative in support of the thesis that power motivation as measured apperceptively reflects a concern about inferiority in general life roles.

Most of the evidence presented suggests that power motivation as measured by the apperceptive device seems to tap concern about weakness or inadequacy, perhaps reflecting what Lazarus (1961) suggests is the defensive substitutive function of fantasy. Other researchers have explicity attempted to get measures that reflect a positive concern. Uleman (1972) and Winter (1971) have tried different research strategies to establish such a measure. Uleman's study attempted to arouse college men's interests in manipulating others by experimental means. He put men in a gambling setting where the subjects could anticipate having high or low control over others. Winter attempted to induce a sense of power by having college men watch pictures of John F. Kennedy—a charisma by association. The scoring schemes for positive power motivation are very different in these two studies. Uleman used a computer-derived conglomeration of criteria—both context and stylistic criteria. Winter mainly focused on imagery relevant to a character who had some impact on the world. None of the scoring systems intercorrelate very highly. Uleman's scoring scheme has little face validity. Winter's scoring scheme seems to be picking up what has been scored for achievement imagery: public recognition of impact, among other things. There is a small positive correlation between power motivation using the Winter scheme and achievement motivation, using the McClelland-Atkinson scheme (1958) (David Winter, personal communication, 1971.) Perhaps the recognition motivational imagery *does* belong within the power domain. Indeed, further work on it might help researchers purify the achievement scoring. This positive orientation to power motivation as measured by either Winter's or Uleman's scheme has had notable success in a number of unpublished studies. However, Winter has established a revised scoring that includes coding from the original Veroff measure, Uleman's and Winter's. Winter also distinguished *positive* and *negative power imagery* in this revised coding procedure. Such a distinction seems important in the future research, one that will take advantage of the cumulative information on the measure discussed in this paper as well as of the information on the measures reflecting more positive orientations to power.

Undoubtedly, there are instances when the power motive even considered as an avoidant motive (as we seem to find with the measure under consideration) can lead to pleasure of accomplishment rather than merely relief from the avoidance of weakness. Where there is a negative motivation, according to both an Adlerian or a positive growth model of motivation, mentioned earlier, there must be some original positive power concerns. Thus, if the conditions are just right, the avoidant concerns might be minimized, power motivation could be freely and pleasurably expressed, and people scoring high on the measure can thereby experience competence in manipulating people. Standards of excellence can be applied to this social skill, as it can to any performance. In fact, we have suggested that power motivation may lead to effective or successful social behaviors, but only under certain circumstances—when the context for their performance is one that is not publicly and clearly defined as power oriented. Do we have any evidence for all of this in the analysis of the current measure of power motivation in interaction with different life roles?

Our analyses of the relationship between power motivation and felt adjustment suggest that power motivation does lead to certain kinds of "positive" social gratifications in role reactions. We found in our analysis that there were instances when power motivation tended to look like the type of motivation that underlies altruistic reactions, but, again, only in roles where there is no public display of the person's power or dominance. Whenever a person's weakness or power can be publicly and accurately gauged by others, then this sort of altruistic reaction based on power does not seem to occur.

We find power motivation underlying apparently altruistic social interests in certain reactions to parenthood and in certain reactions to marriage. In considering the marital role the one result that stood out in this regard was the strong relationship found between power motivation and the mention of relationship satisfactions as the nicest thing about one's marriage. This relationship existed for grade-school-educated women. We found that in this group, in contrast to all other groups, a woman was likely to mention the love relationship with her spouse as being the outstanding feature of her marriage *if* she had high power motivation. Lower status marriages generally have been characterized by *desired* male dominance, but actual feminine control (Zelditch, 1964). This discrepancy allows the wife to find gratification from displaying power on the one hand and not being defensive about her submissive status on the other hand. In her role the grade-school-educated woman can feel comfortable about her sense of weakness and turn to the marital relationship itself for what she can get from it. Her power is not open to public display. A college-educated woman in her marriage is expected to have greater freedom of decision making and hence is more open to public scrutiny. She and the world can ask about how much power she can wield.

With regard to the findings concerning parenthood, we find that high power-motivated young fathers have a positive orientation to their fatherhood. They seem to enjoy it. Maybe this is because they are obliged to perform certain socialization functions in their role. They have to influence their children; they have to transmit norms. But what is special in this case is that the influencing is very private. The public does not evaluate how weak or strong an influencer a father is when his children are very young. It is as if a father with high power motivation can relax in considering his potential weakness when he is in interaction with a very young child. He can turn his social motivations into constructive influencing functions and hence feel very pleased about the role he is in.

Some further enlightenment about how being a father to young children may be positive for a power-motivated father comes from an intensive study of parents and their reported behavior with their nursery-school-age children. Veroff[8] showed that men with high power motivation reported being very affiliative with their nursery-school-age children. Hoffman (1960), using the same data, found a positive relationship between power motivation in fathers and the assertive behavior of their wives toward their nursery-school-age children. It does not seem too farfetched to suggest that men who are concerned about weakness may marry assertive, dominant women who forcefully fulfill role expectations in controlling and socializing young children. In such a situation fathers may first of all be relieved of any responsibility to exercise power themselves in relation to the young children; they may respond affiliatively toward the young child in an attempt to compensate for the mother's harshness; they may secondarily be able to influence the young child by means of their affiliative behavior.

These instances of power motivation being translated into socially desirable and personally fulfilling gratifications in social roles seem to us to be examples of positive outcomes of power motivation. We refer to them in a discussion of the development of social motivation (Veroff & Veroff, 1971).

## POWER MOTIVATION AND SOCIAL ADJUSTMENT

In the nationwide sample survey we also assessed affiliation motivation (the concern about attaining or maintaining a positive emotional relationship with another person) and achievement motivation (the concern about competition with a standard of excellence), as well as indices of felt adjustment based on factor analysis (see Veroff, Feld, & Gurin, 1962). We found that the group of subjects highest in felt adjustment were men and women who were high in all three motives we assessed—achievement, affiliation, and power. It is as if all of these motivations are requisite for adequate adjustment to society. If only one of these motives is dominant, then there are some difficulties encountered. At different ages different motivational patterns seem to define the most distressed group, but typically it was the pattern of one motive high while the others were low that was associated with the most felt difficulty among people. For example, among middle-aged men, those who were low in achievement and affiliation motives but high in power motivation felt particularly socially inadequate. This same pattern for middle-aged women was associated with physical and psychological symptoms. In all age groups men who had high power motives but low affiliation and achievement motives were likely to come from broken home backgrounds (divorce, separation, or death in the family before the person was 16). Of the 65 men who showed the pattern of low achievement and affiliation motivation combined with high power motivation, 29% came from broken home backgrounds of one sort or another, whereas this was true for only 16% of men who showed other motivational combinations. Terhune (1968) recently showed that men who have this pattern are those who are most exploitative in the prisoner dilemma game. They do not take into consideration the needs of a partner.

---

[8] J. B. Veroff. An exploratory study of parental motives, parental attitudes, and social behavior of children. Unpublished doctoral dissertation, University of Michigan, 1958.

Successful adjustment to our complex society seems to require differentiated motivational investments. Recent work by McClelland, Davis, Kalin, and Weimer (1971) suggests that power motivation, as measured through fantasy, is highly correlated with alcoholism. Their interpretation is that fear of inadequacy and impotence underlies alcoholism more than the often-mentioned system of dependency. This may be another instance of lack of adjustment occurring with an overriding motivational theme in a person's life.

In light of this interpretation it would be instructive to look a little more closely at those people whose motivational investments *are* more complex—those who were high in all three motives we assessed. While the men of this type primarily fall into the college-educated group, many of them come from *lower status backgrounds* in contrast to the usual high-status background of most college-educated men. It is interesting to speculate that the high power motivation of these men has, indeed, evolved from their initial feelings of inadequacy, which stimulated efforts to overcome obstacles to acquire status in the society. Perhaps the combination of a strong need to do well, a sensitivity to social demands, and a willingness to do battle to overcome obstacles is the basic equipment of the "self-made man," who is then eminently well-qualified to handle the multiple role demands of our complex society without undue conflict or discontent. This line of reasoning suggests that the current black power movement in this country will be successful if inferiority concerns are amalgamated with achievement and/or social interests.

A differentiated pattern of motivation like the one discussed here may allow for the development of positive power motivation that we have found so difficult to assess. Perhaps people who are high in power motivation but also high in other social motives do not get defensive about power; they may not overinterpret social situations in power terms, which may lead to debilitating social performances, as Berlew suggested. While the original condition for its development may be concern over weakness or inadequacy, perhaps the felicitous combination with other motives allows for the development of a disposition for active enjoyment from the exercise of influence. For example, Browning and Jacobs (1964) reported results about recruitment of politicians which suggest that when power and achievement motivation are both high, then businessmen are likely to get into political office. The most consistent result in looking at both political activity of businessmen in an eastern seaboard city and in a Louisiana setting was, however, the finding that high achievement motivation correlated with seeking and being in a political office with high influence potential. These results suggest that power motivation might not be the most striking determinant of overt power strivings, but when power motivation occurs in conjunction with an achievement orientation, then it can be an important predictor of clear power behavior.

Winter's (1968) scoring system for positive power motivation seems to be a blend of achievement and power strivings. With this blend the goal of power begins to lose its interpersonal quality and takes on the quality of competence. Manipulating people can be like manipulating objects. Under these circumstances one might think that standards of excellence might be applied to how well a person overcomes the difficulty of exerting power. Atkinson has made risk taking about the evaluations of one's performance the core quality of the achievement goal. It might also be applicable to the power setting.

## SUMMARY

We have seen that there are certain regularities in the way the power motive, sometimes in interaction with other motives, seems to be associated with feelings of adjustment in specific and general life roles. To summarize thus far, we make the following statements about high power motivation:

1. It occurs in status groups that are concerned about their weakness;
2. It is correlated with positive social performance and adjustment when the power demands are not publicly salient;
3. It can lead to the avoidance of the power situation, including self-destruction;
4. It can engender conflict if it is the only motivational theme in a person's life style;
5. It can be associated with a differentiated zestful life in combination with other motives (affiliation and achievement).

From the summarized data about the measure we have some very important guidelines to help in deriving a theory of power motivation that encompasses these assertions. We are able to measure a very salient *fear of weakness,* one that seems to be instigated by a disruption in status or access to decision making. This fear can lead to (*a*) avoidant reactions in some instances, (*b*) some costs in personal adjustment in other instances, and (*c*) very positive adjustment in still other instances. A theory of power motivation, notes for which are presented elsewhere (Veroff & Veroff, 1971), should thus specify two different but possibly related conceptual schemes for handling this conclusion. First, there ought to be a developmental theory that specifies three types of characterological power motivation: one that generally leads to avoidant responses, one that leads to costly successful adjustment, and one that leads to very positive adjustment. Second, there should be an ahistorical theory of power motivation that specifies the conditions under which these three types of reactions to instigated power motivation are apparent in behavior, regardless of the power disposition of the individual. Needless to say, a complex theory of power motivation would further specify the interaction between power dispositions and situational power determinants.

## REFERENCES

Adler, A. *The individual psychology of Alfred Adler.* New York: Basic Books, 1956.

Andrews, J. D. W. The achievement motive and advancement in two types of organizations. *Journal of Personality and Social Psychology,* 1967, **6**, 163–168.

Atkinson, J. W. *Motives in fantasy action and society.* Princeton: Van Nostrand, 1958.

Atkinson, J. W. *An introduction to motivation.* Princeton: Van Nostrand, 1964.

Barrett, G. V., & Franke, R. H. Psychogenic death: A reappraisal. *Science,* 1970, **167**, 304–306.

Berlew, D. E. Interpersonal sensitivity and motive strength. *Journal of Abnormal and Social Psychology,* 1961, **63**, 390–394.

Browning, R. P., & Jacobs, H. Power motivation and the political personality. *Public Opinion Quarterly,* 1964, **28**, 75–90.

de Charms, R. *Personal causation.* New York: Academic Press, 1968.

Feld, S. C. Longitudinal study of the origins of achievement strivings. *Journal of Personality and Social Psychology,* 1967, **7**, 408–414.

Feld, S. C., & Smith, C. An analysis of the objectivity of the method of content analysis. In J. W. Atkinson (Ed.), *Motives in fantasy action and society.* Princeton: Van Nostrand, 1958.

Goodman, L. A., & Kruskal, W. Measures of association for cross-classifications. *Journal of the American Statistical Association,* 1954, **49**, 732–764.

Hoffman, M. L. Power assertion by the parent and its impact on the child. *Child Development,* 1960. **31**, 129–144.

Johnson, T. J., & Smith, L. M. Achievement affiliation and power motivation in adolescents. *Psychological Reports,* 1965, **16**, 1249–1252.

Lazarus, R. S. A substitutive defensive conception of apperceptive fantasy. In J. Kagan & G. S. Lesser (Eds.), *Contemporary issues in thematic apperceptive methods.* Springfield, Ill.: Charles C Thomas, 1961.

Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. Ronald Press, 1944.

McClelland, D. C. *The achieving society.* Princeton: Van Nostrand, 1961.

McClelland, D. C., Davis, W. N., Kalin, R., & Weimer, H. Eric. *Alcohol and human motivation.* Chicago: Free Press, 1971.

McKeachie, W. J. Motivation teaching methods and college learning. *Nebraska Symposium on Motivation,* 1961, **9**, 111–146.

Minton, H. L. Power as a personality construct. In B. A. Maher (Ed.), *Progress in experimental personality research.* Vol. 4. New York: Academic Press, 1967.

Rudin, S. A. The personal price of national gl *Trans-action,* 1965, **2**, 4–9.

Schopler, J. Social Power. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 2. New York: Academic Press, 1965.

Skolnick, A. Motivational imagery and behavior over twenty years. *Journal of Consulting Psychology,* 1966, **30**, 463–478.

Smith, M. B. Competence and socialization. In J. Clausen (Ed.), *Socialization and society.* Boston: Little Brown, 1968.

Terhune, K. W. Motives, situation, and interpersonal conflict within prisoner's dilemma. *Journal of Personality and Social Psychology,* 1968, **8,** (Monograph Suppl. 3, Part 2).

Tolman, E. C. Principles of performance. *Psychological Review,* 1955, **62,** 315–326.

Uleman, J. S. The need for influence: Development and validation of a measure of comparison with the need for power. *Genetic Psychology Monograph,* 1972, **85,** 157–214.

Veroff, J. Development and validation of projective measure of power motivation. *Journal of Abnormal and Social Psychology,* 1957, **54,** 1–9.

Veroff, J. A scoring manual for the power motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society.* Princeton: Van Nostrand, 1958.

Veroff, J., Atkinson, J. W., Feld, S. C., & Gurin, G. The use of thematic apperception to assess motivation in a nationwide interview study. *Psychological Monographs,* 1960, **74**(12, Whole No. 499).

Veroff, J., & Feld, S. C. *Marriage and work in America.* New York: Van Nostrand-Reinhold, 1970.

Veroff, J., Feld, S. C., & Gurin, G. Dimensions of subjective adjustment. *Journal of Abnormal and Social Psychology,* 1962, **64,** 192–205.

Veroff, J., & Veroff, J. B. Theoretical notes on power motivation. *Merrill Palmer Quarterly,* 1971, **1,** 59–69.

Weems, L. B., & Wolowitz, H. M. The relevance of power themes among male negro and white paranoid and non-paranoid schizophrenics. *International Journal of Social Psychiatry,* 1969, **15,** 189–196.

Weinstein, M. B. Achievement motivation and risk preference. *Journal of Personality and Social Psychology,* 1969, **13,** 153–172.

White, R. W. Motivation reconsidered: The concept of competence. *Psychological Review,* 1959, **66,** 297–333.

Winter, D. G. The need for power in college men. In D. C. McClelland, W. N. Davis, R. Kalin, H. Weimer, & A. Eric (Eds.), *Alcohol and human motivation.* Chicago: Free Press, 1971.

Wolowitz, H. M., & Shorkey, C. Power themes in the TAT stories of paranoid schizophrenic males. *Journal of Projective Techniques and Personality Assessment,* 1966, **30,** 591–596.

Zelditch, M. Family, marriage and kinship. In R. E. L. Farris (Ed.), *Handbook of modern sociology.* Chicago: Rand McNally, 1964.

(Received February 24, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 4, 292–310

# SEPTUM AND BEHAVIOR:

## A REVIEW[1]

P. A. FRIED[2]

*Carleton University, Ottawa, Canada*

Research and discussion related to the behavioral effects of septal dysfunction are reviewed. It is concluded that a unitary conceptualization of the role of the septum is inappropiate and simplistic as discrete lesion and stimulating techniques indicate that the septum cannot be considered equipotential in either its contribution to various types of behavior or in its anatomical relationships with other subcortical structures. It is proposed that various nuclei in the septum are important in at least four dissociable behavior patterns: (*a*) damage to the medial septal nucleus impairs the ability to alter a response on the basis of proprioceptive cues; (*b*) damage to the lateral and posterior regions of the septum increases the magnitude of negative responses to unpalatable solutions; (*c*) damage to regions of the septum that are anatomically connected to the hypothalamus increases the intake of water; and (*d*) damage to a variety of septal nuclei results in overresponding to positively motivating stimuli. This latter finding is interpreted as implicating the septum in a specific type of response-suppressing mechanism.

In recent years considerable research has been directed toward determining the role of the septum. The multiplicity of experiments has implicated the septal region in a wide variety of behaviors, but an orderly pattern of function has not emerged. It is the purpose of this paper to provide a review of the literature pertaining to septal function and to categorize the behavioral changes associated with damage to the structure.

The septal nuclei are anterior midline structures located in the telencephalon ventral to the anterior portion of the corpus callosum with many connections to both telencephalic and diencephalic structures—particularly the hippocampus and the hypothalamus (e.g., Raisman, Cowan, & Powell, 1966). A detailed report of the neuroanatomy of the septal region is given elsewhere (Fried, 1972). It is not surprising that experimental intervention with the septum by such means as stimulation and lesioning has been found to affect a wide range of behaviors.

## MOTIVATIONAL CHANGES

Probably the best known effect of septal damage is the transformation of a docile rat into one that readily attacks objects brought into its environment, displays explosive startle reactions, and vigorously resists handling. "Septal rage" or hyperreactivity in the rat was reported by Brady and Nauta (1953, 1955) and has been observed by many subsequent workers. Much research has been directed at determining the stimulus conditions and anatomical structures that attenuate or release "septal rage," but neither the essential behavioral conditions for its elicitation nor the anatomical and neurochemical substrates of the phenomenon have been clearly determined. Due to the volume of literature on this topic, it is considered in detail in a separate paper (Fried, see Footnote 3). One aspect of the increase in reactivity that is particularly puzzling is that septal hyperreactivity is not correlated with other response measures. The dissociation of septal hyperreactivity from other effects of septal injury supports the suggestion of several investigators that a unitary conceptualization of function of this structure may be inappropriate (Carey, 1967b; Holdstock, 1970; Morgane, 1969; Novick & Pihl, 1969; Schwartzbaum, Green, Beatty, & Thompson, 1967; Van Hoesen, MacDougall, & Mitchell, 1969). Hy-

---

[1] Research leading to the present paper was supported by Grant MA-3430 from the Medical Research Council of Canada and Grant DRB 9425-24 from the Defense Research Board.

[2] The author would like to thank J. Kelly, A. Laver, and W. Webster for their critical comments.

Requests for reprints should be sent to P. A. Fried, Department of Psychology, Carleton University, Ottawa, Ontario, Canada.

poactivity after septal injury occurs in environments that are not conducive to exploratory behavior—such as in running wheels (Anderson, 1970; Clody & Carlton, 1969; Douglas & Raphelson, 1966; Gotsick, 1969; Nielson, McIver, & Boswell, 1965; Rasmussen, Kaada, & Bruland, 1960), in home cages (Thomas, Moore, Harvey, & Hunt, 1959), and in small rectangular enclosures (Zucker, 1965). However, septal-lesioned rats demonstrate hyperactivity when situations, such as mazes, facilitate exploratory behavior (Nielson et al., 1965; Thomas et al., 1959) and when a familiar environment is changed (Corman, Meyer, & Meyer, 1967; Gotsick, 1969). The data derived from open-field tests appear equivocal; Schwartzbaum and Gay (1966) and Corman et al. (1967) reported a decrease in activity of septal-lesioned animals in open-field tests, whereas Donovick and Wakeman (1969) reported a greater amount of activity. This apparent discrepancy, however, may be due to procedural differences. In both studies in which a decrease in activity level was found, the animals were tested preoperatively in the same apparatus, whereas, in Donovick's experiment, the rats were not placed in the open field until after postoperative recovery. Schwartzbaum and Gay (1966) and Corman et al. (1967), when testing for postoperative activity, placed the septal-lesioned animals in a relatively familiar environment that may not have been conducive to exploration. Related to this point is the report of Douglas and Raphelson (1966) that increased exploratory behavior in septal-lesioned animals occurred only in the first trial. These findings on activity levels are closely related to the question of rate of habituation. Investigators have reported that, relative to control subjects, septal-lesioned animals display faster habituation (Douglas & Raphelson, 1966). Gotsick (1969), who examined this behavior in detail, found that the more rapid habituation of rats with septal injury as measured in activity cages was not present under conditions of food or water deprivation or auditory stimulation. These findings relating the interaction of septal damage, activity, and habituation to motivational or arousal factors confirm earlier results and have been extended by other work. Septal-lesioned animals have previously been reported to manifest increased reactivity to mild, novel stimuli such as light (Clody & Carlton, 1969; Donovick, 1968; Donovick & Wakeman, 1969; Green, Beatty, & Schwartzbaum, 1967), and chemical stimuli (Beatty & Schwartzbaum, 1967) and, recently, Anderson (1970) found that increased activity following septal lesions occurred under conditions of food deprivation and, significantly, under stimulus conditions associated with feeding.

Increased activity following septal lesions has also been observed in two-way shuttle box avoidance tasks (Dalby, 1970; Donovick, 1968; Green et al., 1967; Schwartzbaum et al., 1967) when the behavior was measured during the intertrial intervals and during intershock intervals in shock threshold experiments (Lints & Harvey, 1969a; Lubar, Brener, Deagle, Numan, & Clemens, 1970). However, Kenyon and Krieckhaus (1965a) found no increase in activity during the intertrial interval of a shuttle avoidance task in animals with either unilateral or bilateral septal lesions. Although the reason for this discrepancy cannot be established with certainty at this time, it is interesting to note that the typical lesions of this study primarily destroyed the lateral septal nucleus and spared the medial septal nucleus. That this latter structure may be important in increased intertrial activity has been demonstrated by Donovick (1968) who reported that lesions restricted to the medial nucleus and dorsal portions of the diagonal band resulted in greater activity, whereas lesions that spared the anterior two-thirds of the medial nucleus but injured the lateral septal nucleus failed to increase activity. In addition, Donovick found that lesions of the posterior septum, which generally included the septo-fimbrial and triangular nuclei, fornix and fimbria, also resulted in increased intertrial activity. Both the anteromedial portion of the septum and the fornix have efferent connections with the hippocampus, and, as the lateral septal nuclei project primarily to diencephalic and midbrain structures (Raisman, 1966), it appears that the hippocampus is directly involved in this form of behavior. This proposal is supported by Green et al. (1967) who found increased intertrial activity during a shuttle avoidance experiment with hippocampal-lesioned rats. Clody and Carlton (1969) have shown that lesions restricted to the diagonal band of Broca and the medial septal nucleus result in reduced activity as measured in a running wheel. Since hippocampal-lesioned rats either demonstrate no change in activity (Douglas, 1967) or an increase (Gotsick, 1969) in a similar apparatus, it is probable that this behavior reflects the interaction of the medial septum with nonhippocampal structures. One such structure may be the amygdala, since it is anatomically associated with the medial septal nucleus via the diagonal band of Broca (Clody & Carlton, 1969).

The picture that emerges with respect to activity levels following septal lesions in rats is consistent with Gotsick's (1969) proposal that spontaneous activity is, at least partially, mediated throught the septum.

Following septal damage, rats display more sperm plugs than control subjects (Beach, 1966; Clody & Carlton, 1969), although the interpretation of this is not clear. The greater number of sperm plugs may represent a decreased tendency to consume the seminal product. The strength of heterosexual drive, measured in a box that required rats to cross an electrified grid to get to a sex incentive, was found to increase significantly following dorsal septal lesions (Rasmussen et al., 1960). The few studies examining the effects of septal lesions in cats have been inconclusive (Hagamen & Zitzmann, 1959; Hagamen, Zitzmann, & Ungewitter, 1959). In rats (Slotnick, 1967) unlike mice (Carlson & Thomas, 1968) maternal behavior was not disrupted by extensive lesions. The difference between the effects of septal lesions in mice and rats may be due, as proposed by Thomas, Hostetter, and Barker (1968), to a greater degree of "corticalization of function" of maternal behavior in the rat, as compared to the mouse. Another aspect of social behavior that is modified by septal damage is an increase in interanimal aggression (Ahmad & Harvey, 1968; Blanchard & Blanchard, 1968; Brady & Nauta, 1953; Bunnell, Bemporad, & Flesher, 1966; Bunnell & Smith, 1966; Sodetz & Bunnell, 1970).

In a widely cited study, Lints and Harvey (1969a) reported that large lesions of the rat's septum produced a significant decrease in jump thresholds but did not affect flinch thresholds. This would imply that the septal-lesioned animals in Lints and Harvey's (1969a) study were demonstrating greater sensitivity to suprathreshold stimuli but were not exhibiting any change in their detection limen. This interpretation is supported by the recent work of Lubar et al. (1970). One study has been reported in which septal lesions did not result in changed flinch, jump, or vocalization thresholds (Blanchard & Fial, 1968). The reason for this inconsistency is not apparent although a possible factor confounding the results is that the rats had been observed in an inescapable shock situation and had been trained in a passive avoidance task prior to being tested for shock thresholds.

It has been suggested that sensitivity to painful stimuli may be a function of serotonin content within particular regions of the brain and that the neurohumor has the role of attenuating the effect of aversive stimuli (Harvey & Lints, 1971; Lints & Harvey, 1969a, 1969b; Lorens, Sorenson, & Harvey, 1970; Tenen, 1967). The basis for this hypothesis is the finding that lesions to anatomically interrelated structures such as the medial forebrain bundle (Harvey & Lints, 1971; Lints & Harvey, 1969a), dorsal-medial tegmentum (Lints & Harvey, 1969a), the accumbens nuclei (Lorens et al., 1970), and the septum (Lints & Harvey, 1969a) all result in rats demonstrating an increased sensitivity to shock and decreased brain concentrations of serotonin.

## DIETARY CHANGES

Damage to the septum of the rat brain has produced a reliable increase in water intake (Blass & Hanson, 1970; Donovick & Burright, 1968; Donovick, Burright, & Gittelson, 1968; Hamilton, 1970; Harvey & Hunt, 1965; Harvey, Lints, Jacobson, & Hunt, 1965; Kasper-Pandi, Schoel, & Zysman, 1969; Lorens & Kondo, 1969; Lubar, Schaefer, & Wells, 1969; Pizzi & Lorens, 1967;

Singh & Meyer, 1968; Wolfe & Lubar;[3] Wolfe, Lubar, & Ison, 1967). The exact anatomical locus of the effect is not known, although polydipsia has been related to damage of the posterior septal region (Blass & Hanson, 1970; Lubar, Boyce, & Schaefer, 1968; Lubar et al., 1969; Wolfe & Lubar, see Footnote 3; Wolfe, Lubar, & Ison, 1967). However, Besch and Van Dyne (1969) reported that ventral lesions with damage to the periventricular nucleus were associated with water consumption although no particular septal locus was identified, and Carey (1969) found that both anterior and posterior septal lesions increase water intake. It may be that the behavioral effect of localized damage to the posterior septum is a result of interrupted efferent fibers. It is unlikely that irritative factors are the underlying cause of increased drinking because ablation techniques have been reported to produce the phenomenon (Lubar et al., 1969; Wolfe et al., 1967) and continued increased water consumption continues several months after surgery (Harvey & Hunt, 1965) when irritative effects should have dissipated. The mechanism implicated in the upset water balance has been thought to be an interference with the supraoptic and paraventricular nuclei of the hypothalamus which, in turn, interfere with the production or release of antidiuretic hormone (ADH) (Besch & Van Dyne, 1969; Endroczi & Lissak, 1960; Lubar et al., 1969). Because the hypothalamus is essential to the normal regulation of water intake and because the septum is, anatomically, closely associated with the structure (Raisman, 1966, 1969a, 1969b), it would appear reasonable that, as suggested by Mogenson (Mogenson, McLaughlin, Wishart, & Stevenson;[4] Wishart & Mogenson, 1970a, 1970b), the increased drinking following septal damage may arise from the removal of a source of inhibition for the lateral hypothalamus. The reports that electrical stimulation of the septum decreased water intake under ad-libitum (Wishart & Mogenson, 1970a) and deprivation conditions (Mabry & Peeler, 1968; Wishart & Mogenson, 1970a) are consonant with this hypothesis. Enhanced self-stimulation of the lateral hypothalamus following septal lesions (Lorens, 1966) and decreased stimulation thresholds in the lateral hypothalamic nucleus after damage of the septum (Keesey & Powley, 1968) provide further evidence that the lateral hypothalamus is inhibited by the septal region.

Pharmacological studies have shown that the septal area contains a cholinergic system capable of regulating water intake (Fisher & Coury, 1962; Grossman, 1964). This system, however, is not localized, for cholinergic stimulation of a variety of limbic structures in the rat results in similar drinking behavior (Coury, 1967; Fisher, 1969; Levitt, White, & Sander, 1970), and cholinergic blocking agents, such as atropine, when applied to these limbic structures, prevent drinking in response to cholinergic stimulation of other parts in the system (Levitt & Fisher, 1966). Atropine, when applied to the septum, was reported by Grossman (1964) to reduce water consumption in both satiated and thirsty rats. However, Fisher (1969) did not find any effect of atropine on water ingestion following deprivation and this issue has yet to be resolved. Licking responses to sucrose, a positive incentive stimulus, were significantly greater in septal-lesioned rats than control subjects (Beatty & Schwartzbaum, 1967, 1968a), and responses to water diluted with quinine, a negative incentive stimulus, were considerably reduced (Beatty & Schwartzbaum, 1967; Donovick, Burright, & Zurombaum, 1967; Donovick, Burright, & Zuromski, 1970) suggesting that, as with spontaneous activity (discussed later), septal injury results in amplification of both facilitory and inhibitory behavior when fluids are accepted or rejected.

Food intake appears less under the influence of the septum than liquid consumption (Carlson & Cole, 1970; Gurowitz & Lubar, 1966; Zucker, 1965). A transient increase in food intake was observed by Pizzi and Lorens

---

[3] J. W. Wolfe & J. F. Lubar. The effect of medial cortical lesions on water and food ingestion in the albino rat. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, N.J., 1965.

[4] G. J. Mogenson, R. McLaughlin, T. B. Wishart, & J. A. F. Stevenson. Effects of sodium phenobarbital on water intake of rats with lesions of the ventromedial-medial hypothalamus or septum. Paper presented at the meeting of the Federation of American Societies for Experimental Biology, Atlantic City, N. J., 1969.

(1967) and by Lorens and Kondo (1969). The reasons for the transitory increases are not clear but are probably secondary to the increased water intake. Electrical stimulation of the rats' septum has no significant effect on food intake (Mabry & Peeler, 1968; Wishart & Mogenson, 1970a) although food deprivation increased lever-pressing rates for septal self-stimulation (Brady, Boren, Conrad, & Sidman, 1957; Olds, 1958). There has been one report of hyperphagia following injury of the septal region in the rat (Singh & Meyer, 1968) and as in hypothalamic hyperphagia, if the dietary quality of the food is lowered (by adding powdered cellulose), septal-lesioned animals become aphagic relative to the control subjects.

Because adrenergic stimulation of the septum increases food intake significantly (Booth, 1967; Coury, 1967; Fisher, 1969), while cholinergic stimulation results in increments in drinking, Coury (1967) has suggested that the septum may function to balance food and water intake so that an appropriate level of each is maintained. This hypothesis has been proposed as a possible explanation for differences observed between septal-lesioned and intact animals with respect to water intake levels during food deprivation (Wishart & Mogenson, 1970b). Unoperated animals reduce their consumption of water during food deprivation, but animals with septal lesions show the opposite tendency (Singh & Meyer, 1968; Wishart & Mogenson, 1970b). These results are consonant with the hypothesis that septal damage upsets the relationship between food and water intake and that, following injury, the two systems become somewhat independent.

## SCHEDULES OF REINFORCEMENT

Raphelson, Isaacson, and Douglas (1966), using food reinforcement, reported that rats with large lesions ran down the alley faster after septal damage than they did preoperatively, and that the experimental animals displayed more resistance to extinction than control subjects. On the other hand, Wolfe et al. (1967) found that rats with moderately large septal lesions and no preoperative training ran somewhat slower than control subjects during acquisition in a straight alley with water as reinforcement, and were less resistant to extinction. The acquisition data have been confirmed by Carey (1969), who used rats with posterior septal lesions, but he found that animals with anterior septal damage ran slightly faster than control subjects. Because of the differences in procedures in these studies, comparisons and conclusions do not seem possible.

In lever-pressing tasks under continuous reinforcement conditions, the data are not much clearer. Harvey et al. (1965) did not find any increase in lever-pressing output, with water as reinforcement. However, these workers maintained their animals under a rather severe water deprivation schedule and allowed their subjects only 9 minutes per session under continuous reinforcement conditions. Harvey and Hunt (1965) have reported that septal-lesioned animals manifest a greater lever-pressing output only if the water deprivation regimen is not too severe and if each session during continuous reinforcement lasts for at least 16 minutes. Harvey and Hunt attributed the greater lever-pressing output to increased thirst, because preloading the septal-lesioned rats with water abolished the differences between the control and experimental subjects. An increased level of operant responding by septal-lesioned rats has also been reported by Fallon and Donovick (1970) and Lorens and Kondo (1969). In the latter study, ad-libitum food consumption was also measured, with no increase being found in the septal-lesioned animals. The authors interpreted the increased operant responding with no corresponding ad-libitum food intake as indicating that septal damage increases the sensitivity of rats to drive stimuli developed during deprivation. This interpretation of the effects of septal damage is consonant with the report that septal-lesioned rats increase bar pressing to a greater degree than control animals when maintained in a cold environment (Wakeman, Donovick, & Burright, 1970) and with the finding that rats with septal damage respond significantly less in a lever-pressing task during extinction under water deprivation conditions if they have acquired the response under food deprivation (Fallon & Donovick, 1970).

Numerous investigators have examined the effect of septal damage under fixed-interval reinforcement schedules. Rats with septal lesions exhibit higher response rates than control animals in lever-pressing tasks under water deprivation conditions (Ellen & Powell, 1962a, 1962b; Harvey & Hunt, 1965), under food deprivation conditions (Beatty & Schwartzbaum, 1968b; Lorens & Kondo, 1969; Pubols, 1966; Schwartzbaum & Gay, 1966), and under nondeprivation conditions (Beatty & Schwartzbaum, 1968b). These data, and the report that changes in palatability of the reinforcement do not interact with the lesion effects in this situation (Pubols, 1966), indicate that the dysfunction produced by septal lesions under a fixed-interval reinforcement schedule can be dissociated to a considerable degree from the consummatory changes that follow septal damage described earlier. Furthermore, the elevation in response rates by septal-damaged rats does not appear to be a general impairment in the ability to suppress a response, because the increased bar pressing occurs primarily near the beginning of each fixed interval early in training and then shifts to the late portion of each fixed interval after several days of training (Beatty & Schwartzbaum, 1968b; Ellen & Powell, 1962a, 1962b; Schwartzbaum & Gay, 1966). The data have been interpreted by some as demonstrating that the septal damage results in a specific disinhibition of responses related to motivationally significant stimuli (Beatty & Schwartzbaum, 1968b; Ellen, 1966) while other researchers have suggested that septal damage results in increased response perseveration when appropriate proprioceptive cues are prominent (McCleary, 1966).

In addition to the fixed-interval schedule, several other patterns of intermittent reinforcement have been used with animals with septal damage. Ellen and Powell (1962a) found that neither the acquisition nor the retention of a lever-pressing response in rats with medial septal lesions under a fixed-ratio schedule of water reinforcement differed from those of control subjects. However, Carey (1969), who also used water in a fixed-ratio schedule, reported that rats with large anterior septal lesions emitted significantly more responses than control subjects during the acquisition of a lever-pressing response, while animals with large posterior septal lesions pressed the lever significantly fewer times during training. When the same animals were tested under a fixed-ratio food reinforcement regimen, none of the groups differed significantly from one another. The findings were interpreted by Carey as reflecting a combined dysfunction, following septal damage, of increased thirst and increased perseveration. As Carey's anterior and posterior septal lesions result in opposite behaviors under a fixed-ratio water reinforcement schedule, and as Ellen and Powell's (1962a) rats had septal damage that overlapped the two regions lesioned by Carey, the differences between the studies may have an anatomical basis.

Generally speaking, the most consistent picture to emerge from studies examining intermittent reinforcement regimens is that septal damage affects responding during interval schedules to a greater extent than during ratio schedules.

If reinforcement is contingent on the animal withholding its response for a specific time interval, the mode of the characteristic frequency distribution of responses in intact animals is at approximately the minimum interresponse interval required for reinforcement. This schedule is called the differential reinforcement of low response rates (DRL) and has been used extensively to examine timing behavior of septal-lesioned animals. Ellen, Wilson, and Powell (1964) reported that small, medial septal lesions in rats disrupted a food-motivated, preoperatively established DRL performance, with the dysfunction being manifested by a marked increase in emitted responses occurring after short interresponse intervals. The anatomical locus within the septum which, when lesioned, results in an increased rate of responding during the DRL schedule, appears to be the ventral portion of the medial septal nucleus and the diagonal band region, under either food (Burkett & Bunnell, 1966) or water (Carey, 1968b) reinforcement conditions.

The impaired ability of septal-lesioned rats in a DRL task has been shown by Caplan and Stamm (1967) to depend, in part, on the training procedures used. The frequency distribution of interresponse times of the animals with septal damage does not differ from that of the control subjects if the DRL temporal contingency is lengthened in gradual, discrete steps. This finding may account for the failure of Harvey and Hunt (1965) to detect any impairment in performance of septal-lesioned animals in a DRL task in which the required interresponse interval was gradually increased.

It is unlikely that the disruption of performance on a DRL task by septal-lesioned rats can be attributed to an increase in appetitive behavior, since lever pressing for water on a DRL schedule and water intake can be independently increased by discrete septal lesions (Carey, 1967b). Furthermore, Carey (1967c) has shown that the distribution of interresponse times of intact, water-deprived rats on a DRL schedule is markedly different from that generated by septal-lesioned rats. It is also unlikely that the deficit in DRL performance is a symptom of impaired motor inhibition or of an impaired timing mechanism as the increase in responding primarily occurs early in the interreinforcement interval (Ellen et al., 1964; Thomas et al., 1968). In addition, septal-lesioned rats have been shown to be capable of learning both an easy (short interresponse interval) DRL task (Caplan, 1970; Caplan & Stamm, 1967; Thomas et al., 1968) and a difficult DRL task, when a light served as an exteroceptive cue signaling the end of the required delay (Ellen & Butter, 1969). In the reports of Caplan (1970) and Thomas et al. (1968), the actual number of reinforcements received by the experimental and control subjects in the easy DRL task was quite similar, although the septal-lesioned rats were making many more bar presses. However, these responses occurred during the early portion of the interreinforcement interval and therefore had a minimal effect in delaying the reinforcement. In Caplan's (1970) study, reinforcement was occasionally withheld following a correctly timed response. In such an instance, the septal-lesioned rats increased their rates of responding significantly more than that of control subjects, suggesting some sort of enhanced frustration effect (Amsel, 1958; 1962). This report complements the findings of Simmons (Thomas et al., 1968) that septal-lesioned rats tend to emit, in an easy DRL situation, more short-latency responses relative to control animals after a nonreinforced response but not after a reinforced response. It may be noteworthy that the hippocampal rhythm present in response to frustrative nonreward has the lowest threshold for septal driving of all hippocampal theta activity (Gray & Ball, 1970).

On a DRL schedule, an animal must rely on response-produced proprioceptive stimuli to determine whether or not an adequate duration has elapsed since the last response. Ellen and Powell (1966) found that if electrical stimulation of the septum was used to indicate the end of the required interresponse interval, rats learned a DRL task quite readily. Subsequent to that study, Ellen and Butter (1969) reported that septal-lesioned rats could perform a DRL task if the end of the interreinforcement period was signaled by a light. The ability of septal-lesioned animals to perform the DRL task in these two studies was attributed to the exteroceptive cues of brain stimulation and light which could serve as discriminative cues for bar pressing. In the more typical DRL task, because there are no exteroceptive stimuli present to signal the end of the delay, only response-produced feedback stimuli are available; septal damage is thought to impair the ability to use proprioceptive stimuli as discriminative cues.

## LEARNING

Septum dysfunction has been shown to have differential effects in a wide variety of discrimination problems. The establishment of a position habit in rats is not disrupted either by stimulation of the lateral septal nuclei (Kasper, 1965) and the anteromedial region of the septum (Donovick & Schwartzbaum, 1966), or by extensive septal lesions (Hamilton, 1970; Hamilton, Kelsey, & Grossman, 1970; Schwartzbaum & Donovick, 1968), or by smaller lesions damaging the medial and lateral nuclei selectively (Donovick, 1968; Hamilton et al., 1970). The ability to acquire a position habit also remains unimpaired in cats with dorsal septal lesions (Zucker & McCleary, 1964). The reversal of a postoperatively acquired position habit, however, is disrupted in rats by stimulation of either the lateral (Kasper, 1965) or medial septal nuclei (Donovick & Schwartzbaum, 1966) and by lesions that cause extensive septal damage (Hamilton, 1970; Hamilton et al., 1970; Schwartzbaum & Donovick, 1968). Rats with lesions that spared most of the medial septal nucleus were less deficient in the position reversal task than animals with these nuclei damaged (Hamilton et al., 1970). Rats with unilateral damage to the lateral nucleus were intermediate to control subjects and rats with bilateral damage on the reversal task (Hamilton, 1970). In cats, the ability to reverse a position habit was impaired by dorsal septal lesions although, when they were required to learn a second reversal, there were no differences between the experimental and control groups (Zucker & McCleary, 1964). The decrease of lesion effects with repeated practice in a position reversal task has also been found in rats with medial septal damage (Donovick, 1968).

Septal dysfunction affects neither the acquisition of a simultaneous discrimination task in rats (Donovick & Schwartzbaum, 1966; Liss & Lukaszewska, 1966; Schwartzbaum & Donovick, 1968; Schwartzbaum, Kellicutt, Spieth, & Thompson, 1964) and cats (Zucker, 1965), nor the reversal of such a discrimination in rats (Donovick & Schwartzbaum, 1966; Schwartzbaum & Donovick, 1968) and cats (Zucker, 1965) and mice (Slotnick & Brown, 1970). Schwartzbaum et al. (1964) found that even though the total acquisition scores were similar for septal-lesioned and control rats in a simultaneous discrimination task, there was an increase in perseverative errors in the early postoperative sessions with the level of responding to the reinforced stimulus not changing but the number of responses during the presentation of the nonreinforced tone and during extinction increasing. These results are consistent with the view described earlier that septal lesions may result in an enhanced frustration effect.

A variety of alternation tasks have been used to examine the effects of septal lesions. Large lesions in the rat have been reported to block the acquisition of a spatial alternation task (Schwartzbaum & Donovick, 1968) and subsequent work (Donovick, 1968) has confirmed an earlier finding (Thompson & Langer, 1963) that destruction of the medial septal nucleus may be responsible for the dysfunction. Stimulation of the medial septal nuclei and the diagonal band system did not impair the acquisition of a spatial alternation task (Schwartzbaum & Donovick, 1968). In mice, destruction of the septal region was found to enhance the learning of a spatial alternation problem (Carlson & Cole, 1970). In monkeys, lesions of the medial and lateral septal nuclei did not impair the retention of a preoperatively established delayed alternation problem (Butters & Rosvald, 1968). This study may be confounded as several other problems were interposed between surgery and postoperative testing on the delayed alternation task. In a nonspatial alternation task, in which reinforcement was given on alternate trials for running down a straight alley, septal-lesioned rats did not develop the normal alternation in running speeds (Thomas et al., 1968). The difficulty septal-lesioned animals have in this task and in the reversal of a postoperatively established position habit contrasts with the lack of difficulty such animals have in simultaneous discrimination reversals. This disparity is in accord with the hypothesis that perseveration (or heightened frustration effect) may be due to septal damage resulting in an inability to modify responses to postoperatively acquired proprioceptive cues (Ellen & Butter, 1969; McCleary, 1966). One study that appears to contradict this proposal is the report of Dalland (1970) that, in a T maze in which the start-stem portion could be rotated 180 degrees, septal-lesioned rats, on the second trial, perseverated toward the stimuli in the maze arm rather than persisting in making the same bodily response turn. It is possible, however, that one trial in the maze was not sufficient to develop a proprioceptive tendency, and what was being observed was the heightened exploratory tendency discussed previously in this review.

There have been surprisingly few studies examining the effects of septal damage in mazes. In a study that supports the proprioceptive hypothesis mentioned above, septal-lesioned rats acquired a food-reinforced kinesthetic discrimination, but were impaired in the subsequent reversal problem (Gittelson & Donovick, 1968). In a more complex maze (Lashley Type III), septal lesions were reported to result in slightly more errors only in those subjects who displayed septal hyperirritability (Thomas et al., 1959). The authors concluded that the minor disruptive effect of septal damage was attributable to an increase in persistent exploratory behavior. Carey (1968a) found that large lesions of the rat's septum resulted in an almost complete retention loss of a preoperatively trained multiple-maze habit. The lack of any savings was not correlated with emotionality and was attributed by Carey to an impairment in memory.

Studies examining the effects of septal damage on the acquisition of a one-way avoidance task are quite varied in their results. Some authors have reported impaired learning (Kenyon & Krieckhaus, 1965b; McNew & Thompson, 1966; Vanderwolf, 1964); others have found no change (Dalby, 1970; Hamilton, 1969; Lubar, 1964; Zucker & McCleary, 1964); and still others report improved learning (Zucker, 1965). McCleary (1966) has proposed that deficits in one-way tasks by septal-damaged animals are attributable to the handling of the subjects between successive trials. Support for this contention is derived from the fact that the only experiment performed with septal-lesioned rats in which no impairment was found in one-way avoidance used an apparatus that did not require handling of the animals between trials (Dalby, 1970). Moreover, all but one of the studies (Hamilton, 1969) that did not find an acquisition impairment in septal-lesioned cats used a nonhandling avoidance task.

Retention of a preoperatively acquired avoidance habit is impaired after septal lesions in cats if a nonhandling procedure is used (McCleary, Jones, & Ursin, 1965). Septal-lesioned cats did not differ from control animals when the extinction procedure involved the punishment of the previously correct response (Zucker, 1965; Zucker & McCleary, 1964). When the extinction paradigm is modified so as to include a novel conditioned stimulus, increased resistance to extinction has been found in lesioned cats (Zucker, 1965) and in cats receiving anticholinergic stimulation (Hamilton, McCleary, & Grossman, 1968). Hamilton (1969), however, failed to find any difference between cats with small septal lesions and control animals in a similar extinction situation.

The faster acquisition of a two-way avoidance response by septal-lesioned rats, first reported by King (1958), has been found by many workers (Donovick, 1968; Garber & Simmons, 1968; Green et al., 1967; Green & Schwartzbaum, 1968; Hamilton, 1969, 1970; Kasper-Pandi et al., 1969; Kenyon & Krieckhaus, 1965a; Krieckhaus, Simmons, Thomas, & Kenyon, 1964; Meyer, Johnson, & Vaughn, 1970; Novick & Pihl, 1969; Schwartzbaum et al., 1967; Trafton, 1967; Usher & Lamble, 1969; Van Hoesen et al., 1969). The same phenomenon has been reported for mice (Carlson, 1970) and for cats (Fox, Kimble, & Lickey, 1964; McCleary, 1961). Injections of cholinergic blocking agents in the septum of cats duplicated the effect of septal lesions in this task (Hamilton & Grossman, 1969), and also improved performance of a previously acquired avoidance response (Grossman, 1964). A study by Moore (1964) was inconsistent with this latter finding; he reported that lesions of the cat's posterior septum resulted in impaired retention of a preoperatively learned avoidance response. Although the enhanced learning is a consistent finding, with only three contrary reports (Dalby, 1970; LaVaque, 1966; Kasper-Pandi et al., 1969), there is little agreement on the interpretation of the effects of septal damage in this task. It is unlikely that, as proposed by King (1958), the facilitated two-way learning can be attributed to hyperemotionality, because several authors have found that increased irritability can be disassociated from the improved acquisition of the avoidance response (Krieckhaus et al., 1964; Schwartzbaum et al., 1967; Trafton, 1967). An alternative hypothesis for the enhanced learning is that septal damage reduces freezing responses (the tendency to crouch), and so, in the two-way avoidance problem, the conflict between the tendency to flee from the shock compartment and the tendency to not enter a previously shocked area (i.e., to freeze) is more easily resolved in the lesioned animals (Donovick, 1968; Garber & Simmons, 1968; Kenyon & Krieckhaus, 1965a; Krieckhaus et al., 1964; Schwartzbaum et al., 1967). This interpretation is supported by the report of Garber and Simmons (1968) that, in comparison to control subjects, the septal-lesioned rats were superior in escaping from the low-shock to the high-shock side of a shuttle box but were inferior in performing the response in the reverse direction when the conflict was at a minimum. Furthermore, the investigators who failed to find more rapid acquisition following septal damage used either a very low shock level (Kasper-Pandi et al., 1969; LaVaque, 1966) or a very brief shock presented intermittently with a long shock–shock interval (Dalby, 1970). Some results, however, do not support the decreased freezing hypothesis. The facilitating effect of septal lesions was found to be independent of shock level when rats were trained under either high-shock or low-shock conditions (Schwartzbaum et al., 1967; Thomas et al., 1968), although in neither study was the low-shock level as low as the one used by Garber and Simmons (1968), Kasper-Pandi et al. (1969), or LaVaque (1966). Data that also appear inconsistent with the freezing hypothesis include both the Meyer et al. (1970) finding that the enhanced learning by septal-lesioned rats was not entirely attributable to increased intertrial responding, and the reports of various passive-avoidance studies in which rats with septal damage learned the task in a normal fashion (Fried, 1969; Gittelson, Donovick, & Burright, 1969). Further, as discussed earlier in this report, and as McCleary (1966) has indicated, perseveration of approach responses in septal-lesioned animals occurs in a variety of tasks where the freezing response does not normally occur.

A somewhat more general hypothesis has been proposed by McCleary (1961, 1966). He has interpreted the two-way avoidance data and other behavioral results discussed above as indicating that one consequence of septal damage is a lack of response inhibition. However, it is apparent from many of the studies considered in this review that disinhibition is not an inevitable consequence of injury to the septum. Several authors (Gerbrandt, 1965; Hamilton, 1970; McCleary, 1966) have theorized that the loss of response suppression results in a release of those behaviors that are highest in habit strength for a particular situation, thereby preventing new, and possibly more appropriate, responses from becoming "stabilized" (Gerbrandt, 1965). This view is consistent with the increased resistance to extinction of a two-way conditioned avoidance response found in septal-lesioned animals (Hamilton, 1970; Kasper-Pandi et al., 1969; LaVaque, 1966; Van Hoesen et al., 1969), and is considered further in the review of the passive avoidance literature.

No particular anatomical region within the septum has been associated with facilitated learning of the two-way avoidance problem (Donovick, 1968). Unilateral lesions (Green & Schwartzbaum, 1968; Hamilton, 1970; Hamilton et al., 1970; Kenyon & Krieckhaus, 1965a), unilateral stimulation (Gardner & Malmo, 1969; Goldstein, 1966a), and small lesions (Kenyon & Krieckhaus, 1965a) produce effects at least as large as those found after bilateral lesions.

A variation of the two-way active avoidance situation is the Sidman avoidance task (Sidman, 1953). In this paradigm a subject receives a shock if its response, usually moving from one compartment to another in a shuttle box, is withheld longer than a time interval determined by the experimenter. The most efficient method of responding is for the subject to withhold his response until just prior to the end of the intershock interval. Septal-lesioned rats learned a Sidman schedule more rapidly than controls and also performed at a closer to optimal level (Sodetz, 1970). This finding provides additional support to the hypothesis that response disinhibition is not an inevitable consequence of septal damage.

A few studies have investigated the behavior of septal-lesioned animals in active avoidance tasks without using various types of shuttling responses as dependent measures. Cats with septal lesions (Hamilton, 1969) or with cholinergic blocking agents injected into the septum (Hamilton & Grossman, 1969) did not readily learn to jump onto a shelf to avoid shock, although short-latency escape responses quickly developed. These results were interpreted as being consistent with the hypothesis, mentioned earlier, that effects of septal damage in active avoidance situations are dependent upon whether the appropriate avoidance response is high on the response hierarchy.

Nielson et al. (1965) used an avoidance task in which shock could be terminated by pressing the correct lever in one of the arms of a T maze. The arm selected as the incorrect choice was that chosen spontaneously by rats on the first trial. Septal-lesioned animals were deficient in acquiring this avoidance response. McCleary (1966) has suggested that this deficit may be considered as an example of response perseveration, manifested by impaired ability to reverse an original position habit.

As in the one-way shuttle avoidance task (McCleary et al., 1965) preoperatively acquired active avoidance responses are impaired following septal damage. The retention of presurgically established responses of bar pressing to escape noise (Tracy & Harrison, 1956) and of clinging to a wall to escape shock (Rich & Thompson, 1965) was not found in septal-lesioned rats. Both of these studies, as well as the experiment of McCleary et al. (1965), are characterized by the fact that no intertrial handling occurred. This may be an important variable in determining the effects of septal damage (McCleary, 1966), but no data have been reported on the retentive abilities of septal-lesioned animals in active avoidance tasks in which handling was present.

One of the first experimental manipulations performed with septal-damaged rats used the conditioned emotional response (CER) paradigm. With intact animals, the basic procedure for the CER task usually consists of establishing a stable, positively reinforced lever-pressing response and then presenting a conditioned stimulus for a specified interval, followed by a foot shock. After a few such pairings, a reduction in the number of lever presses occurs during the conditioned stimulus. This reduction, the CER, is thought to represent "anxiety" arising from the anticipation of an inescapable shock (Hunt & Brady, 1951). The effect of septal lesions on the acquisition of a CER has not been reliably established. Impairments in septal-lesioned rats have been reported by Trafton (1967) who tested the animals 10 days after surgery using a high-shock level, and by Harvey, Jacobson, and Hunt (1961) and Harvey et al. (1965), using a low-shock level, when training began more than 3 months after surgery. No deficits in acquisition have been found by Brady and Nauta (1953) with a high-shock level and with training beginning 3 days after surgery, by Harvey et al. (1965) who used a high-shock level either 48 days or 184 days after surgery, and by Harvey et al. (1961), who used a high- or low-shock level 48 days after surgery. Septal self-stimulation impairs the development of a CER and also interferes with a previously established CER to a water-reinforced habit (Brady & Conrad, 1960). As noncontingent rewarding septal stimulation also interferes with the suppression of a water-motivated lever press (Goldstein, 1966b), it would appear that the disruption of the CER is not due to an incentive mechanism.

The retention of a preoperatively acquired CER is impaired following septal lesions, and the impairment appears to be independent of shock level or time of testing (Brady & Nauta, 1953, 1955; Harvey et al., 1961; Harvey et al., 1965).

Another type of response suppression paradigm extensively used to study septal-lesioned animals is that in which subjects are punished for making a particular response. The ability to learn this passive avoidance task has been found, in many studies, to be impaired following septal dysfunction (Fox et al., 1964; Fried, 1969, 1970, 1971; Gurowitz & Lubar, 1966; Hamilton, 1970; Hamilton et al., 1970; Hamilton, McCleary, & Grossman, 1968; Harvey et al., 1965; Kaada, Rasmussen, & Kviem, 1962; Kasper, 1964; Kasper-Pandi et al., 1969; Lubar, 1964; McCleary, 1961; McCleary et al., 1965; McGowan, Garcia, Ervin, & Schwartz, 1969; McNew & Thompson, 1966; Middaugh & Lubar, 1970; Schwartzbaum & Spieth, 1964; Slotnick & Brown, 1970; Slotnick & Jarvik, 1966; Thomas et al., 1968; Ursin, Linck, & McCleary, 1969; Van Hoesen et al., 1969; Wishart & Mogenson, 1970c; Zucker, 1965; Zucker & McCleary, 1964). However, many

authors have reported that, under certain circumstances, passive avoidance performance is not retarded in septal-lesioned animals. One variable that appears important in determining whether or not impaired passive avoidance performance occurs is the strength of the response to be suppressed. In situations in which the response to be inhibited was not previously reinforced, septal-lesioned rats did not differ from control subjects in postshock passive avoidance (Blanchard & Fial, 1968; Middaugh & Lubar, 1970; Winocur & Mills, 1969). When the habit strength of a to-be-suppressed response was experimentally manipulated by varying the number of prepunishment training trials, the passive avoidance deficit in septal-lesioned subjects was positively correlated with the number of preshock trials (Middaugh & Lubar, 1970; Schwartzbaum & Spieth, 1964). These data are consistent with the hypothesis mentioned above that the loss of response suppression results in the release of behaviors that have been previously stabilized (Gerbrandt, 1965). An extension to this hypothesis has been recently proposed. Fried (1969, 1970, 1971) has observed passive avoidance deficits in septal-lesioned animals only when postoperative approach training preceded the shock trials. Preoperative training did not result in postoperative approach perseveration except when approach training was also given after surgery. The results indicated that in passive avoidance situations the behavior of septal-lesioned rats is independent of the high habit strength responses present prior to surgery but is dependent upon the approach responses developed after surgery. This is consistent with the previously cited passive avoidance studies which found deficits as, in all of those reports, septal lesions were made prior to various types of approach training. It is also consonant with the report (Carey, 1967a) that resistance to extinction was increased in septal-lesioned animals if a response was acquired after surgery but decreased if it was acquired prior to surgery. On the basis of these results and much of the data reviewed above, it is hypothesized that, as Zucker (1965) had previously suggested, one of the effects of septal damage is to increase to a greater than normal degree the response habit that follows an actual or expected reinforcement.

As well as the strength of the approach response in a passive avoidance conflict, the intensity and the type of the aversive stimulus have been found to be important variables, although the results of the various studies are not consistent. Harvey et al. (1965) found significant passive avoidance deficits in a lever-pressing task only when low shocks were used, whereas Schwartzbaum and Spieth (1964), using the same task and the same strain of rats, found that the differences between septal-lesioned and control animals increased with increasing shock. Thomas et al. (1968), using a running response, reported that there was a rather restricted range of shock levels that could be used to differentiate septal-lesioned and control rats, and that the shock level interacted with the strain of rat being used. Several studies suggest that the type of aversive stimulus used may be an important variable. McGowan et al. (1969) reported that the pairing of flavored water with nausea-producing X-ray exposure did not result in a passive avoidance deficit in septal-lesioned rats, and Gittelson et al. (1969) found that when quinine was used as an aversive stimulus, septal-lesioned animals were facilitated in the acquisition of a passive avoidance problem.

Low-level septal stimulation facilitated an escape reaction to peripheral foot shock but, when the aversive stimulus was central stimulation of the dorsomedial tegmentum, septal stimulation either inhibited or facilitated escape behavior, depending upon whether the septal stimulation, at higher intensities, was negative or positively reinforcing, respectively (Gardner & Malmo, 1969; Routtenberg & Olds, 1963). It was suggested that the differential interaction of septal stimulation and central or peripheral aversive stimulation was due to "action systems" (Melzack & Wall, 1965) being triggered in different ways as varied anatomical pathways were utilized (Gardner & Malmo, 1969). This theorizing may also be applicable to the studies of McGowan et al. (1969) and Gittelson et al. (1969).

Few investigators have specified anatomically the neural structures within the septum which, when damaged, result in passive avoidance deficits. Hamilton et al. (1970) found that lateral septal-lesioned rats were more impaired in a passive avoidance task than rats with medial septal lesions, and similar results have been reported by Kasper (1964) with stimulating techniques.

## CONCLUSIONS

It is apparent that the septal area is involved in several fundamental and dissociable aspects of behavior. The varied effects of septal dysfunction are probably due to the multiple anatomical connections the septum has with subcortical regions, particularly the hypothalamus and hippocampus. Discrete lesion and stimulating techniques have shown that the septum cannot be considered equipotential in its contribution to various types of behavior, and, therefore, in those studies in which extensive septal damage was produced, the resulting behavior probably represents the combination of differential multiple effects.

One behavior pattern in which the septum plays an important role is the ability to alter a response on the basis of proprioceptive cues. This was evident in such seemingly diverse tasks as position reversals, successive discrimination reversals, and various schedules of reinforcement. The evidence from a number of studies (e.g., Burkett & Bunnell, 1966; Donovick, 1968; Hamilton et al., 1970) indicates that this dysfunction is correlated with damage to the medial septal nucleus. As damage to the hippocampus (Beatty & Schwartzbaum, 1968b; Lash, 1964) or the fornix (McCleary, 1966) results in similar behavioral alterations and as the medial septal nucleus has efferent and afferent connections to the hippocampus via the fimbria and dorsal fornix (Raisman, 1969a, 1969b), it is likely that this aspect of septal dysfunction is mediated via the septal-hippocampal system.

A second behavior related to septal dysfunction but not associated with the one described above is an increased aversion to unpalatable solutions (Donovick et al., 1970; Gittelson et al., 1969). Damage restricted to the anterior portion of the septum does not result in this motivational change (Donovick et al., 1970), whereas lesions of the entire septum (Donovick et al., 1970), the lateral and posterior septal nuclei (Donovick et al., 1970), and the habenular nucleus (Donovick, Burright, Kaplan, & Rosenstreich, 1969) do increase rejection of quinine. One might infer that this aspect of behavior is mediated by efferents originating in the lateral septum and septofimbrial nucleus and projecting via the stria medullaris to the habenular nucleus (Nauta, 1960; Raisman, 1966).

The septum also plays a role in the regulation of water and, possibly, other palatable fluids as an increase in intake follows septal damage or cholinergic stimulation. This behavior is not related to the behaviors previously described (Beatty & Schwartzbaum, 1968a; Carey, 1967b; Donovick et al., 1960). The specific anatomical loci within the septum mediating this change have not been determined, although it is likely that the effect is due to inhibition of the lateral hypothalamus by portions of the septum.

One of the most widely attributed roles of the septum is that of functioning in a response-suppression system (McCleary, 1966). On the basis of a number of studies considered in this review, a more specific hypothesis seems possible. It is proposed that the septum is involved in dampening the habit strength of responses developed during motivating situations, thereby allowing new responses to become established. Animals with septal damage overrespond to motivating stimuli in a variety of situations, increase their rate of operant responding if positive reinforcement is received or expected, and are impaired in their ability to suppress responses that have been positively reinforced in the past. As others have indicated, it is apparent that, following septal damage, the loss of inhibition resulting in a release of stabilized responses is operating at a rather complex level (Ellen, 1966; Gerbrandt, 1965; McCleary, 1966; Ursin et al., 1969). That is, the response-suppressing system, of which the septum is hypothesized to be part, is not operating at the level of individual muscle tonicity, but rather at the higher level of complex, organized, goal-directed responses. Although the anatomical locus of the behavior cannot be identified at this time, it is reasonable to assume that septal nuclei and fibers that are involved include ones that are functionally not connected with the hippocampus, as damage to hippocampal afferents from the septum and to the hippocampus itself results in different behavioral dysfunctions than damage to the septal region when tested in particular passive avoidance tasks (Donovick, 1968; Fried, 1969, 1970, 1971; Fried & Goddard, 1967; McCleary, 1966; Van Hoesen et al., 1969) and under particular reinforcement schedules (Ellen, 1966; Ellen & Powell, 1962b).

## REFERENCES

AHMAD, S. S., & HARVEY, J. A. Long-term effects of septal lesions and social experience of shock-elicited fighting in rats. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 596–602.

AMSEL, R. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review*, 1962, **69**, 306–328.

ANDERSON, R. A. Appetitively motivated general activity in rats with limbic lesions. *Physiology and Behaviour*, 1970, **5**, 755–761.

BEACH, F. A. Sexual behavior in the male rat. *Science*, 1966, **153**, 769–770.

BEATTY, W. W., & SCHWARTZBAUM, J. S. Enhanced reactivity to quinine and saccharine following septal lesions in the rat. *Psychonomic Science*, 1967, **8**, 483–484.

BEATTY, W. W., & SCHWARTZBAUM, J. S. Consummatory behavior for sucrose following septal lesions in the rat. *Journal of Comparative and Physiological Psychology*, 1968, **65**, 93–102. (a)

BEATTY, W. W., & SCHWARTZBAUM, J. S. Commonality and specificity of behavioral dysfunctions following septal and hippocampal lesions in rats. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 60–68. (b)

BESCH, N. F., & VAN DYNE, G. C. Effects of locus and size of septal lesion on consummatory behaviour in the rat. *Physiology and Behaviour*, 1969, **4**, 953–958.

BLANCHARD, R. J., & BLANCHARD, D. C. Limbic lesions and reflexive fighting. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 603–605.

BLANCHARD, R. J., & FIAL, R. A. Effects of limbic lesions on passive avoidance and reactivity to shock. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 606–612.

BLASS, E. M., & HANSON, D. G. Primary hyperdipsia in the rat following septal lesions. *Journal of Comparative and Physiological Psychology*, 1970, **70**, 87–93.

BOOTH, D. A. Localization of the adrenergic feeding system in the rat diencephalon. *Science*, 1967, **158**, 515–517.

BRADY, J. V., BOREN, J. J., CONRAD, D. G., & SIDMAN, M. The effect of food and water deprivation upon intracranial self-stimulation. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 134–137.

BRADY, J. V., & CONRAD, D. G. Some effects of limbic system self-stimulation upon conditioned emotional behavior. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 128–137.

BRADY, J. V., & NAUTA, W. J. H. Subcortical mechanisms in emotional behavior: Affective changes following septal forebrain lesions in the albino rat. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 339–346.

BRADY, J. V., & NAUTA, W. J. H. Subcortical mechanisms in emotional behavior: The duration of affective changes following septal and habenular lesions in the albino rat. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 412–420.

BUNNELL, B. N., BEMPORAD, N. J., & FLESHER, C. F. Septal forebrain lesions and social dominance behavior in the hooded rat. *Psychonomic Science*, 1966, **6**, 207–208.

BUNNELL, B. N., & SMITH, M. Septal lesions and aggressiveness in the cotton rat. *Psychonomic Science*, 1966, **6**, 443–444.

BURKETT, E. E., & BUNNELL, B. N. Septal lesions and the retention of DRL performance in the rat. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 468–471.

BUTTERS, N., & ROSVALD, H. E. Effect of septal lesions on resistance to extinction and delayed alternation in monkeys. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 389–395.

CAPLAN, M. Effects of withheld reinforcement on timing behavior of rats with limbic lesions. *Journal of Comparative and Physiological Psychology*, 1970, **71**, 119–135.

CAPLAN, M., & STAMM, J. DRL acquisition in rats with septal lesions. *Psychonomic Science*, 1967, **8**, 5–6.

CAREY, R. J. A retention loss following septal ablations in the rat. *Psychonomic Science*, 1967, **7**, 307–308. (a)

CAREY, R. J. Independence of effects of septal ablations on water intake and response inhibition. *Psychonomic Science*, 1967, **8**, 3–4. (b)

CAREY, R. J. Contrasting effects of increased thirst and septal ablations on DRL responding in rats. *Physiology and Behaviour*, 1967, **2**, 287–290. (c)

CAREY, R. J. Impairment of maze retention resulting from septal injury. *Physiology and Behaviour*, 1968, **3**, 495–497. (a)

CAREY, R. J. A further localization of inhibitory deficits resulting from septal ablation. *Physiology and Behaviour*, 1968, **3**, 645–649. (b)

CAREY, R. J. Contrasting effects of anterior and posterior septal injury on thirst motivated behaviour. *Physiology and Behaviour*, 1969, **4**, 759–764.

CARLSON, N. R. Two-way avoidance behavior of mice with limbic lesions. *Journal of Comparative and Physiological Psychology*, 1970, **70**, 73–78.

CARLSON, N. R., & COLE, J. R. Enhanced alternation performance following septal lesions in mice. *Journal of Comparative and Physiological Psychology,* 1970, **73**, 157–161.

CARLSON, N. R., & THOMAS, G. J. Maternal behavior of mice with limbic lesions. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 731–737.

CLODY, D. E., & CARLTON, P. L. Behavioral effects of lesions of the medial septum of rats. *Journal of Comparative and Physiological Psychology,* 1969, **67**, 344–351.

CORMAN, C. D., MEYER, P. M., & MEYER, D. R. Open-field activity and exploration in rats with septal and amygdaloid lesions. *Brain Research,* 1967, **5**, 469–476.

COURY, J. N. Neural correlates of food and water intake in the rat. *Science,* 1967, **156**, 1763–1765.

DALBY, D. A. Effect of septal lesions in the acquisition of two types of active avoidance behavior in rats. *Journal of Comparative and Physiological Psychology,* 1970, **73**, 278–283.

DALLAND, T. Response and stimulus perseveration in rats with septal and dorsal hippocampal lesions. *Journal of Comparative and Physiological Psychology,* 1970, **71**, 114–118.

DONOVICK, P. J. Effects of localized septal lesions on hippocampal EEG activity and behavior in rats. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 569–578.

DONOVICK, P. J., & BURRIGHT, R. G. Water consumption of rats with septal lesions following two days of water deprivation. *Physiology and Behaviour,* 1968, **3**, 285–288.

DONOVICK, P. J., BURRIGHT, R. G., & GITTELSON, P. L. The effects of septal lesions on saccharine choice as a function of water deprivation. *Physiology and Behaviour,* 1968, **3**, 677–681.

DONOVICK, P. D., BURRIGHT, R. G., KAPLAN, J., & ROSENSTREICH, N. Habenular lesions, water consumption and palatability of fluids in the rat. *Physiology and Behaviour,* 1969, **4**, 45–47.

DONOVICK, P. J., BURRIGHT, R. G., & ZUROMSKI, E. Localization of quinine aversion within the septum, habenula, and interpeduncular nucleus of the rat. *Journal of Comparative and Physiological Psychology,* 1970, **71**, 376–383.

DONOVICK, P. D., & SCHWARTZBAUM, J. S. Effects of low-level stimulation of the septal area on two types of discrimination reversal in the rat. *Psychonomic Science,* 1966, **6**, 3–4.

DONOVICK, P. J., & WAKEMAN, K. A. Open-field luminance and "septal hyper-emotionality." *Animal Behaviour,* 1969, **17**, 186–190.

DOUGLAS, R. J. The hippocampus and behavior. *Psychological Bulletin,* 1967, **67**, 416–442.

DOUGLAS, R. J., & RAPHELSON, A. C. Septal lesions and activity. *Journal of Comparative and Physiological Psychology,* 1966, **62**, 465–467.

ELLEN, P. Some conceptual issues in the analysis of limbic lesions. *Ammon's Horn.,* 1966, **1**, 27–41.

ELLEN, P., & BUTTER, J. External cue control of DRL performance in rats with septal lesions. *Physiology and Behaviour,* 1969, **4**, 1–6.

ELLEN, P., & POWELL, E. W. Effects of septal lesions on behavior generated by positive reinforcement. *Experimental Neurology,* 1962, **6**, 1–11. (a)

ELLEN, P., & POWELL, E. W. Temporal discrimination in rats with rhinencephalic lesions. *Experimental Neurology,* 1962, **6**, 538–547. (b)

ELLEN, P., & POWELL, E. W. Differential conditioning of septum and hippocampus. *Experimental Neurology,* 1966, **16**, 162–171.

ELLEN, P., WILSON, A., & POWELL, E. Septal inhibition and timing behavior in the rat. *Experimental Neurology,* 1964, **10**, 120–132.

ENDROCZI, E., & LISSAK, K. The role of the mesencephalon, diencephalon, and archicortex in the activation and inhibition of the pituitary-adrenocortical system. *Acta Physiologica Hungaria,* 1960, **17**, 39–55.

FALLON, D., & DONOVICK, P. J. Low resistance to extinction in rats with septal lesions under inappropriate appetitive motivation. *Journal of Comparative and Physiological Psychology,* 1970, **73**, 150–156.

FISHER, A. E. The role of limbic structures in the central regulation of feeding and drinking behavior. *Annals of the New York Academy of Sciences,* 1969, **157**(2), 894–901.

FISHER, A. E., & COURY, J. N. Cholinergic tracing of a central neural circuit underlying the thirst drive. *Science,* 1962, **138**, 691–693.

FOX, S. S., KIMBLE, D. P., & LICKEY, M. E. Comparison of caudate nucleus and septal-area lesions on two types of avoidance behavior. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 380–386.

FRIED, P. A. Effects of septal lesions on conflict resolution in rats. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 375–380.

FRIED, P. A. Pre- and post-operative approach training and conflict resolution by septal and hippocampal lesioned rats. *Physiology and Behaviour,* 1970, **5**, 975–979.

FRIED, P. A. Limbic system lesions in rats: Differential effects in an approach-avoidance task. *Journal of Comparative and Physiological Psychology,* 1971, **74**, 349–353.

FRIED, P. A. The septum and hyperreactivity. *British Journal of Psychology,* 1972, in press.

FRIED, P. A., & GODDARD, G. V. The effects of hippocampal lesions at different stages of conflict in the rat. *Physiology and Behaviour,* 1967, **2**, 325–330.

GARBER, E. G., & SIMMONS, H. J. Facilitation of two-way avoidance performance by septal lesions in rats. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 559–562.

GARDNER, L., & MALMO, R. B. Effects of low-level septal stimulation on escape: Significance for limbic-midbrain interactions in pain. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 65–73.

GERBRANDT, L. Neural systems of response release and control. *Psychological Bulletin,* 1965, **64**, 113–123.

Gittelson, P. L., & Donovick, P. J. The effects of septal lesions on the learning and reversal of a kinesthetic discriminaiton. *Psychonomic Science,* 1968, **13**, 137–138.

Gittelson, P. L., Donovick, P. J., & Burright, R. G. Facilitation of passive avoidance acquisition in rats with septal lesions. *Psychonomic Science,* 1969, **17**, 292–293.

Goldstein, R. Facilitation of active avoidance behaviour by reinforcing septal stimulation in the rat. *Physiology and Behaviour,* 1966, **1**, 335–339. (a)

Goldstein, R. Effects of noncontingent septal stimulation of the CER in the rat. *Journal of Comparative and Physiological Psychology,* 1966, **61**, 132–135. (b)

Gotsick, J. E. Factors affecting spontaneous activity in rats with limbic system lesions. *Physiology and Behaviour,* 1969, **4**, 587–593.

Gray, J. A., & Ball, G. G. Frequency-specific relation between hippocampal theta rhythm, behavior, and amobarbital action. *Science,* 1970, **168**, 1246–1248.

Green, R. H., Beatty, W. W., & Schwartzbaum, J. S. Comparative effects of septohippocampal and caudate lesions on avoidance behavior in rats. *Journal of Comparative and Physiological Psychology,* 1967, **64**, 444–452.

Green, R. H., & Schwartzbaum, J. S. Effects of unilateral septal lesions on avoidance behavior, discrimination reversal, and hippocampal EEG. *Journal of Comparative and Physiological Psychology,* 1968, **65**, 388–396.

Grossman, S. P. Effect of chemical stimulation of the septal area on motivation. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 194–200.

Gurowitz, E. M., & Lubar, J. F. Changes in activity, food ingestion, and passive avoidance behavior following limbic-septal ablation in the cat. *Proceedings of the 74th Annual Convention of the American Psychological Association,* 1966, **1**, 107–108. (Summary)

Hagamen, W. D., & Zitzmann, E. K. "Hypersexual" activity in normal male cats. *Anatomical Record,* 1959, **133**, 388.

Hagamen, W. D., Zitzmann, E. K., & Ungewitter, L. H. Sexual activity of male cats following localized lesions of the frontal lobes. *Anatomical Record,* 1959, **133**, 388.

Hamilton, L. W. Active avoidance impairment following septal lesions in cats. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 420–431.

Hamilton, L. W. Behavioural effects of unilateral and bilateral septal lesions in rats. *Physiology and Behaviour,* 1970, **5**, 855–859.

Hamilton, L. W., & Grossman, S. P. Behavioral changes following disruption of central cholinergic pathways. *Journal of Comparative and Physiological Psychology,* 1969, **69**, 76–82.

Hamilton, L. W., Kelsey, J. E., & Grossman, S. P. Variation in behavioral inhibition following different septal lesions in rats. *Journal of Comparative and Physiological Psychology,* 1970, **70**, 79–86.

Hamilton, L. W., McCleary, R. A., & Grossman, S. P. Behavioral effects of cholinergic septal blockade in the cat. *Journal of Comparative and Physiological Psychology,* 1968, **66**, 563–568.

Harvey, J. A., & Hunt, H. F. Effect of septal lesions on thirst in the rat as indicated by water consumption and operant responding for water reward. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 49–56.

Harvey, J. A., Jacobson, L. E., & Hunt, H. F. Long-term effects of lesions in the septal forebrain on acquisition and retention of conditioned fear. *American Psychologist,* 1961, **16**, 449. (Abstract)

Harvey, J. A., & Lints, C. E. Lesions in the medial forebrain bundle: Relationships between pain sensitivity and telencephalic content of serotonin. *Journal of Comparative and Physiological Psychology,* 1971, **74**, 28–36.

Harvey, J. A., Lints, C. E., Jacobson, L. E., & Hunt, H. F. Effects of lesions in the septal area on conditioned fear and discriminated instrumental punishment in the albino rat. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 37–48.

Holdstock, T. L. Plasticity of autonomic functions in rats with septal lesions. *Neuropsychologia,* 1970, **8**, 147–160.

Hunt, H. F., & Brady, J. V. Some effects of electroconvulsive shock on a conditioned emotional response ("anxiety"). *Journal of Comparative and Physiological Psychology,* 1951, **44**, 88–98.

Kaada, B., Rasmussen, E., & Kviem, O. Impaired acquisition of passive avoidance behavior by subcallosal, septal, hypothalamic, and insular lesion in rats. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 661–670.

Kasper, P. Attenuation of passive avoidance by continuous septal stimulation. *Psychonomic Science,* 1964, **1**, 219–220.

Kasper, P. Disruption of position habit reversal by septal stimulation. *Psychonomic Science,* 1965, **3**, 111–112.

Kasper-Pandi, P., Schoel, W. M., & Zysman, M. Motivation and response strength in passive avoidance deficits of septal lesioned rat. *Physiology and Behaviour,* 1969, **4**, 815–821.

Keesey, R. E., & Powley, T. L. Enhanced lateral hypothalamic reward sensitivity following septal lesions in the rat. *Physiology and Behaviour,* 1968, **3**, 557–562.

Kenyon, J., & Krieckhaus, E. E. Enhanced avoidance behavior following septal lesions in the rat as a function of lesion size and spontaneous activity. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 466–469. (a)

Kenyon, J., & Krieckhaus, E. E. Decrements in one-way avoidance learning following septal lesions. *Psychonomic Science,* 1965, **3**, 113–114. (b)

King, F. A. Effects of septal and amygdaloid lesions on emotional behavior and conditioned avoidance responses in the rat. *Journal of Nervous and Mental Disease*, 1958, **126**, 57–63.

Krieckhaus, E. E., Simmons, H. J., Thomas, G. J., & Kenyon, J. Septal lesions enhance shock avoidance behavior in the rat. *Experimental Neurology*, 1964, **9**, 107–113.

Lash, L. Response discriminability and the hippocampus. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 251–256.

LaVaque, T. J. Conditioned avoidance response perseveration in septal rats during massed extinction trials. *Psychonomic Science*, 1966, **5**, 409–411.

Levitt, R. A., & Fisher, A. E. Anticholinergic blockade of centrally induced thirst. *Science*, 1966, **154**, 520–522.

Levitt, R. A., White, C. S., & Sander, D. M. Dose-response analysis of carbachol-elicited drinking in the rat limbic system. *Journal of Comparative and Physiological Psychology*, 1970, **72**, 345–350.

Lints, C. E., & Harvey, J. A. Altered sensitivity to footshock and decreased brain content of serotonin following brain lesions in the rat. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 23–32. (a)

Lints, C. E., & Harvey, J. A. Drug induced reversal of brain damage in the rat. *Physiology and Behaviour*, 1969, **4**, 29–31. (b)

Liss, P., & Lukaszewska, I. The effects of overtraining and septal lesions on the ability to switch attention between cues. *Acta Biologiae Experimentalis*, 1966, **26**, 299–307.

Lorens, S. A. Effects of lesions in the central nervous system on lateral hypothalamic self-stimulation in the rat. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 256–262.

Lorens, S. A., & Kondo, C. Y. Effects of septal lesions on food and water intake and operant responding for food. *Physiology and Behaviour*, 1969, **4**, 729–732.

Lorens, S. A., Sorenson, J. P., & Harvey, J. A. Lesions in the nucleus accumbens septi of the rat. *Journal of Comparative and Physiological Psychology*, 1970, **73**, 284–290.

Lubar, J. F. Effect of medial cortical lesions on the avoidance behavior of the cat. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 38–46.

Lubar, J. F., Boyce, B. A., & Schaefer, C. F. Etiology of polydipsia and polyuria in rats with septal lesions. *Physiology and Behaviour*, 1968, **3**, 289–292.

Lubar, J. F., Brener, J. M., Deagle, J. H., Numan, R., & Clemens, W. J. Effect of septal lesions on detection threshold and unconditioned response to shock. *Physiology and Behaviour*, 1970, **5**, 459–463.

Lubar, J. F., Schaefer, C. F., & Wells, D. G. The role of the septal area in the regulation of water intake and associated motivational behavior. *Annals of the New York Academy of Sciences*, 1969, **157**, 875–893.

Mabry, P. D., & Peeler, D. F. Response rate for food and water in the rat as a function of non-contingent, reinforcing septal stimulation. *Psychonomic Science*, 1968, **13**, 51–52.

McCleary, R. A. Response specificity in the behavioral effects of limbic system lesions in the cat. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 605–613.

McCleary, R. A. Response-modulating functions of the limbic system: Initiation and suppression. In E. Stellar & J. M. Sprague (Eds.), *Progress in physiological psychology*. Vol. 1. New York: Academic Press, 1966.

McCleary, R. A., Jones, C., & Ursin, H. Avoidance and retention deficits in septal cats. *Psychonomic Science*, 1965, **2**, 85–86.

McGowan, B. K., Garcia, J., Ervin, F. R., & Schwartz, J. Effects of septal lesions on bait-shyness in the rat. *Physiology and Behaviour*, 1969, **4**, 907–909.

McNew, J., & Thompson, R. Role of the limbic system in active and passive avoidance conditioning in the rat. *Journal of Comparative and Physiological Psychology*, 1966, **61**, 173–180.

Melzack, R., & Wall, P. Pain mechanisms: A new theory. *Science*, 1965, **150**, 971–979.

Meyer, P. M., Johnson, D. A., & Vaughn, D. W. The consequences of septal and neocortical ablations upon learning a two-way conditioned avoidance response. *Brain Research*, 1970, **22**, 113–120.

Middaugh, L. D., & Lubar, J. F. Interaction of septal lesions and experience on the suppression of punished responses. *Physiology and Behaviour*, 1970, **5**, 233–237.

Moore, R. Y. Effects of some rhinencephalic lesions on retention of conditioned avoidance behavior in cats. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 65–71.

Morgane, P. J. The function of the limbic and rhinic forebrain-limbic midbrain systems and reticular formation in the the regulation of food and water intake. *Annals of the New York Academy of Sciences*, 1969, **157**, 806–848.

Nauta, W. J. H. Some neural pathways related to the limbic system. In E. R. Ramey & D. S. Doherty (Eds.), *Electrical studies of the unanaesthetized brain*. New York: Hoeber, 1960.

Nielson, H., McIver, H., & Boswell, R. Effect of septal lesions on learning, emotionality, activity, and exploratory behavior in rats. *Experimental Neurology*, 1965, **11**, 147–157.

Novick, I., & Pihl, R. Effect of amphetamine on the septal syndrome in rats. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 220–225.

Olds, J. Effects of hunger and male sex hormone on self-stimulation of the brain. *Journal of Comparative and Physiological Psychology*, 1958, **51**, 320–324.

Pizzi, W. J., & Lorens, S. A. Effect of lesions in the amygdalo-hippocampo-septal system on food and water intake in the rat. *Psychonomic Science*, 1967, **7**, 187–188.

Pubols, L. M. Changes in food motivation behaviour in rats as a function of septal and amygdaloid lesions. *Experimental Neurology*, 1966, **15**, 240–254.

Raisman, G. The connexions of the septum. *Brain*, 1966, **89**, 317–349.

Raisman, G. Comparison of the mode of termination of the hippocampal and hypothalamic afferents to the septal nuclei as revealed by electron microscopy of degeneration. *Experimental Brain Research*, 1969, **7**, 317–343. (a)

Raisman, G. Neuronal plasticity in the septal nuclei of the adult rat. *Brain Research*, 1969, **14**, 25–48. (b)

Raisman, G., Cowan, W. M., & Powell, T. P. S. An experimental analysis of the efferent projection of the hippocampus. *Brain*, 1966, **89**, 83–108.

Raphelson, A. C., Isaacson, R. L., & Douglas, R. J. The effect of limbic damage on the retention and performance of a runway response. *Neuropsychologia*, 1966, **4**, 253–264.

Rasmussen, E. W., Kaada, B. R., & Bruland, H. Effects of neocortical and limbic lesions on the sex drive in rats. *Acta Physiologica Scandinavica*, 1960 (Suppl. 175), 126–127.

Rich, I., & Thompson, R. Role of hippocampus-septal system, thalamus and hypothalamus in avoidance conditioning. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 66–72.

Routtenberg, A., & Olds, J. Attenuation of response to an aversive brain stimulus by concurrent rewarding septal stimulation. *Federation Proceedings*, 1963, **22**, 515. (Abstract)

Schwartzbaum, J. S., & Donovick, P. J. Discrimination reversal and spatial alternation associated with septal and caudate dysfunction in rats. *Journal of Comparative and Physiological Psychology*, 1968, **65**, 83–92.

Schwartzbaum, J. S., & Gay, P. Interacting behavioral effects of septal and amygdaloid lesions in the rat. *Journal of Comparative and Physiological Psychology*, 1966, **61**, 59–65.

Schwartzbaum, J. S., Green, R. H., Beatty, W. W., & Thompson, J. B. Acquisition of avoidance behavior following septal lesions in the rat. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 95–104.

Schwartzbaum, J. S., Kellicutt, M. H., Spieth, T. M., & Thompson, J. B. Effects of septal lesions in rats on response inhibition associated with food reinforced behavior. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 217–224.

Schwartzbaum, J. S., & Spieth, T. M. Analysis of the response-inhibition concept of septal functions in "passive avoidance" behavior. *Psychonomic Science*, 1964, **1**, 145–146.

Sidman, M. Two temporal parameters of the maintenance of avoidance behavior by the white rat. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 253–261.

Singh, D., & Meyer, D. R. Eating and drinking by rats with lesions of the septum and the ventromedial hypothalamus. *Journal of Comparative and Physiological Psychology*, 1968, **65**, 163–166.

Slotnick, B. M. Disturbances of maternal behaviour in the rat following lesions of the cingulate cortex. *Behaviour*, 1967, **29**, 204–236.

Slotnick, B. M., & Brown, D. L. Fear conditioning and passive avoidance in mice with septal lesions. *Physiology and Behaviour*, 1970, **5**, 1255–1259.

Slotnick, B. M., & Jarvik, M. Deficits in passive avoidance and fear conditioning in mice with septal lesions. *Science*, 1966, **154**, 1207–1208.

Sodetz, F. J. Septal ablation and free-operant avoidance behaviour in the rat. *Physiology and Behaviour*, 1970, **5**, 773–777.

Sodetz, F. J., & Bunnell, B. N. Septal ablation and the social behaviour of the golden hamster. *Physiology and Behaviour*, 1970, **5**, 79–88.

Tenen, S. S. The effects of p-chlorophenylalanine, a serotonin depletor, on avoidance acquisition, pain sensitivity and related behaviour in the rat. *Psychopharmacologia*, 1967, **10**, 204–219.

Thomas, G. J., Hostetter, G., & Barker, D. J. Behavior functions of the limbic system. In E. Stellar & J. M. Sprague (Eds.), *Progress in physiological psychology*. Vol. 2. New York: Academic Press, 1968.

Thomas, G. J., Moore, R. Y., Harvey, J. A., & Hunt, H. F. Relations between the behavioral syndrome produced by lesions in the septal region of the forebrain and maze learning of the rat. *Journal of Comparative and Physiological Psychology*, 1959, **52**, 527–532.

Thompson, R., & Langer, S. K. Deficits in position reversal learning following lesions of the limbic system. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 982–995.

Tracy, W. H., & Harrison, J. M. Aversive behavior following lesions of the septal region of the forebrain in the rat. *American Journal of Psychology*, 1956, **69**, 443–447.

Trafton, C. L. Effects of lesions in the septal area and cingulate cortical areas on conditioned suppression of activity and avoidance behavior in rats. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 191–197.

Ursin, H., Linck, P., & McCleary, R. A. Spatial differentiation of avoidance deficit following septal and cingulate lesions. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 74–79.

Usher, D. R., & Lamble, R. W. ACTH synthesis and release in septal-lesioned rats exposed to air shuttle avoidance. *Physiology and Behaviour*, 1969, **4**, 923–927.

Vanderwolf, C. H. Effect of combined medial thalamic and septal lesions on active-avoidance behavior. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 31–37.

Van Hoesen, G. W., MacDougall, J. M., & Mitchell, J. C. Anatomical specificity of septal projections in active and passive avoidance behavior. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 80–89.

Wakeman, K. A., Donovick, P. J., & Burright, R. G. Septal lesions increase bar pressing for heat in animals maintained in the cold. *Physiology and Behaviour,* 1970, **5**, 1193–1195.

Winocur, G., & Mills, J. A. Hippocampus and septum in response inhibition. *Journal of Comparative and Physiological Psychology,* 1969, **67**, 352–357.

Wishart, T. B., & Mogenson, G. J. Reduction of water intake by electrical stimulation of the septal region of the rat brain. *Physiology and Behaviour,* 1970, **5**, 1399–1404. (a)

Wishart, T. B., & Mogenson, G. J. Effects of food deprivation on water intake in rats with septal lesions. *Physiology and Behaviour,* 1970, **5**, 1481–1486. (b)

Wishart, T. B., & Mogenson, G. J. Effects of lesions of the hippocampus and septum before and after passive avoidance training. *Physiology and Behaviour,* 1970, **5**, 31–34. (c)

Wolfe, J. W., Lubar, J. F., & Ison, J. R. Effects of medial cortical lesions on appetitive instrumental conditioning. *Physiology and Behaviour,* 1967, **2**, 239–244.

Zucker, I. Effect of lesions of the septal-limbic area on the behavior of cats. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 344–352.

Zucker, I., & McCleary, R. A. Perseveration in cats. *Psychonomic Science,* 1964, **1**, 387–388.

(Received March 1, 1971)

## ERRATA

In the article "Comment on 'Equivalence of Information in Concept Identification'" by Leonard M. Giambra in the September 1972 issue, lines 4–11 on page 188 should read as follows:

That is, the equivalence-of-information method will work when only Formula 1 = Formula 2. One can easily demonstrate that Formula 2 can equal Formula 3 only when there are two values per dimension.[2] That is, Formula 2 = Formula 3 if and only if $X_i$ is two for *all* values of $i$.

In addition, the second sentence of Footnote 2 should begin as follows:

The case where the number of dimensions of the universe is two results in Formula 2 = Formula 3 for any value of $X_i$ . . . .

*Psychological Bulletin*
1972, Vol. 78, No. 4, 311–320

# MAINTENANCE SCHEDULES AND HUNGER DRIVE:
## AN EXAMINATION OF THE RAT LITERATURE

ROY B. WEINSTOCK[1]

*Mary Washington College*

In the interest of examining the possible relevance of the maintenance schedule variable, studies are reviewed in which the feeding method used to produce hunger drive was a manipulated variable and was not confounded with differences in body weight loss. In addition, a standardized notation system, capable of describing various maintenance procedures, was introduced and used in order to facilitate intra- and interexperiment comparisons. On the basis of the studies reviewed, it is concluded that (*a*) the maintenance schedule variable is a relevant one in the determination of behavior; (*b*) hunger drive is probably determined by some joint function of body weight loss and maintenance schedule; and (*c*) based on data currently available, some form of an adjusted percentage maintenance schedule appears to be the most appropriate for use in the study of rat behavior.

Even the most casual survey of the psychology literature that is devoted to the study of infrahuman behavior will reveal two facts: (*a*) the bulk of animal experimentation has involved, and continues to involve, various strains of the laboratory rat as subject (Beach, 1950; Bitterman, 1965); and (*b*) food deprivation is very extensively employed as the means of inducing the experimental subject to "behave." Moreover, the *joint* presence of these two events looms quite large in the day-to-day affairs of the animal psychologist.

The rationale behind the food restriction operation has to do with the traditional assumption of psychologists that by reducing an animal's access to food, a hypothetical state of "drive" is induced in the organism. Drive is then considered to spur the animal into increasing its activity level in such a way that there is held to exist a generally positive relationship between degree of food restriction and response strength. Various measures have been used to index such changes in response output. For example, Bare and Cicala (1960), Bolles (1965), Capaldi and Robinson (1960), Fallon (1965), Hoyenga and Aeschleman (1969), Moll (1964), Siegel (1961), Verplanck and Hayes (1953), and Williams and Campbell (1961) have all looked at the rat's consumatory[2] behavior as a function of food deprivation. Other investigators have focused on changes in general activity as an index of drive (Campbell, Teghtsoonian, & Williams, 1961; Duda & Bolles, 1963; Olewine, Barrows, & Shock, 1964; Strong, 1957; Treichler & Hall, 1962; Weasner, Finger, & Reid, 1960). Similarly, a considerable amount of data has been collected in order to examine the relationship between hunger drive and exploratory behavior (Adlerstein & Fehrer, 1955; Bolles & de Lorge, 1962a; de Lorge & Bolles, 1961; Hughes, 1968; Woods & Bolles, 1965; Zimbardo & Miller, 1953; Zimbardo & Montgomery, 1957), and between hunger drive and instrumental learning (Bare, 1959; Jensen, 1960; Perin, 1942; Stolurow, 1951; Warner, 1928). Taken as a group, these studies indicate that a generally positive relationship exists between severity of food deprivation and performance output.

The picture that emerges from the above examples is that a wide variety of behavioral indexes have been employed to measure the motivational effects of drive, based on food restriction. Also, because a wide variety of

---

[1] The author is greatly indebted to Robert C. Bolles for reading a preliminary version of this paper and making a number of valuable suggestions.

Requests for reprints should be sent to Roy B. Weinstock, Department of Pyschology, Mary Washington College, Fredericksburg, Virginia 22401.

[2] There should be no ambiguity between the words *consummatory* and *consumatory*. As Young (1961, p. 250) has indicated, "The first term derives from *consummate* and the other from *consume*."

TABLE 1

FEEDING OR MAINTENANCE SCHEDULES GENERATED BY DIFFERENT LEVELS OF TEMPORAL, STANDARD, AND INTAKE VARIABLES

| Temporal | Standard | Intake | | |
|---|---|---|---|---|
| | | Amount (A) | Time (T) | Percentage (P) |
| Regular (R) | Fixed (F) | RFA | RFT | RFP |
| | Adjusted (A) | RAA | RAT | RAP |
| Irregular (I) | Fixed | IFA | IFT | IFP |
| | Adjusted | IAA | IAT | IAP |

deprivation regimens all seem to lead to hunger motivation, it might appear that it makes no difference what the particular nature or character of a food deprivation regimen is, for example, whether the rat is to be allowed access to an unlimited food supply for a specific period of time, or whether it is to be allowed unlimited time to consume a specific amount of food. As a result, very little consideration, other than experimental convenience, is typically given to the selection of a particular feeding procedure.

While not a great deal of attention has been paid to the food deprivation variable, there are sufficient data on this problem to examine the tenability of the regimen-equality assumption. The purpose of this paper is to survey the relevant literature and to evaluate critically this assumption.

## PROPERTIES OF FOOD MAINTENANCE SCHEDULES

There are three major parameters that together serve to determine the character of any particular food deprivation regimen. One is a *temporal* parameter and refers to whether or not an animal's access to food is permitted at the same point in time within an arbitrary time band. Typically, such a time band is marked off in multiples of 12- or 24-hour intervals. Thus, the subject can be allowed to feed on a regular, diurnal cycle, basis (e.g., at each 9:00 A.M. and 9:00 P.M.; at every 2:00 P.M.), or on an irregular (non-12- or non-24-hour) basis (e.g., once every 17 hours; once every 24 hours on the average). A second parameter under experimental control pertains to the *standard* against which the subject's intake pattern is measured. Specifically, a schedule can be either adjusted or fixed. As used herein, any adjusted food maintenance schedule employs an external referent according to which adjustment is made. This typically takes the form of yoking an experimental subject's food intake or percentage body weight to that of a control subject. A fixed schedule, on the other hand, is one that does not utilize control information of any sort. It consists of the experimenter selecting a particular numerical value that specifies the quantity of food the subject is to receive, the amount of free-feeding time the subject is to have out of a fixed block of time, or the percentage of predeprivation ad-libitum weight at which the subject is to be maintained. The third parameter is the particular *intake* criterion that the experimenter imposes in determining the rat's access to food. Thus, the experimenter can modulate either the amount of food to be received by the animal, the rat's free-feeding time, or the percentage body weight at which the rat is to be maintained. The various combinations of each parameter level are represented in Table 1.

The first letter in each cell of the table corresponds to the degree of cyclicity and uniformity in an animal's food maintenance regimen (regular or irregular). The second letter indicates whether or not an external standard is to be used in monitoring the subject's deprivation regimen (fixed or adjusted). The third letter denotes the operation that monitors the quantity of food the subject is to receive at its feeding time (amount, time, or percentage). Thus, RAA is the procedure whereby food, adjusted in amount according to some criterion, is presented consistently once every 12- or 24-hour multiple; RAT indicates a procedure whereby the subject, once every 24 hours, for example, is allowed to consume food for a period of time that has been adjusted by some external criterion or standard; and so on. Examination of the 12 procedures that have been outlined in Table 1 indicates that those variables which conventionally serve to define maintenance techniques are represented therein. Thus, the quantity of food allotted to the subject, amount of consumatory time, and weight loss (specified in absolute or percentage terms) are all factors that are represented by the composition of this matrix.

Three brief points should be made at this time. First, all procedures that have been outlined in Table 1, which includes those commonly used by experimenters in general, are confounded in that by specifying either the amount of food, access time to food, or percentage body weight at which the subject is to be maintained, other maintenance factors vary. For example, if the subject is always fed a specific quantity of food at the same time each day (RFA procedure), then the amount of time that the subject has to consume this food is free to vary, and does. As another example, the RFP procedure calls for the subject to be maintained at a given percentage of its predeprivation ad-libitum body weight. Since there will be fluctuations around this percentage figure, the subject will be fed different amounts of food in order to bring its body weight in line with the desired percentage criterion. However, these confoundings are rather unimportant since the confounding factor does not vary systematically, while the specified component of the maintenance procedure (e.g., food amount, percentage body weight) does.

Second, with respect to the intake variable, it should be pointed out that the percentage schedules represent a class of maintenance methods that is conceptually different from the amount and time techniques. Specifically, the former set of procedures is useful in controlling the consequences of food ingestion, while the latter two are used to monitor attendant food consumption.

The third point involves the classification system presented in Table 1. In discussing the studies presented below, it will be necessary to index the particular amount of food, access time to food, or percentage body weight that has been selected by the experimenter to maintain the various groups of rats used in a study. This will allow for an exact specification of the character of a particular food deprivation regimen. Therefore, "g" is used to index the amount of food in grams received by the subject, "m" and "h" stand, respectively, for minutes and hours of the subject's access time to food per block of hours, and % refers to the percentage body weight at which the subject is maintained or the percentage of a control subject's food intake that the experimental subject will receive as a ration. The only exceptions to these rules will occur when a subject is being maintained on one of the two extreme feeding procedures possible. Thus, a continuous deprivation schedule (i.e., one which calls for the subject never being fed) will be referenced by RFTo. An ad-libitum feeding schedule will be indexed by RFTal. By using the above notation system while examining the results of the studies reviewed, it is hoped that intra- and inter-experiment comparison will be facilitated.

Two criteria were used in selecting the studies reviewed below. First, rat studies were chosen in which, within a given study, comparison had been made between at least two different food deprivation procedures, and not between levels of only one maintenance method.[3] That is, data are presented in which maintenance procedure was a manipulated variable and the experimenter's intent was to examine the equivalence, or lack thereof, of different feeding methods. If two groups are treated alike in all respects except for the maintenance schedule under which they have been kept, and they manifest different behavioral patterns in some current or subsequent test situation, this will be taken as presumptive evidence that maintenance method is a variable that enters significantly into the determination of that behavior. Second, only studies in which body weight loss was not confounded with maintenance schedule were considered. Thus, investigations by Baker (1955), Bolles and de Lorge (1962b), Lawrence and Mason (1955), Mirsky, Looney, and Rosvold (1954), and Ramond, Carlton, and McAllister (1955) are not evaluated since in each case intergroup differences in feeding method covaried with differences in body weight loss.

---

[3] Thus, the work of Eisman (1956) and his coworkers (Eisman, Asimow, & Maltzman, 1956; Eisman, Linton, & Theios, 1960; Eisman, Theios, & Linton, 1961) and Ehrenfreund (1959, 1960, 1966) and his colleagues (Ehrenfreund & Allen, 1966; Ehrenfreund & Badia, 1962) is not considered. Eisman's two-parameter theory of hunger drive involves consideration of the hours of deprivation imposed on the subject during a fixed period of time ($h_t$) and the hours of deprivation in force at the time of testing ($h$). Since these two parameters are temporal ones, and since the converse of $h_t$ is the number of free-feeding hours given during the interval in which $h_t$ is occurring, Eisman's basic experimental manipulation can be conceived of as involving different values of the RFT maintenance technique. With respect to Ehrenfreund and the weight control device he originated, examination of his procedure reveals that his manipulation of hunger drive involves maintaining animals at various levels of the IFP procedure.

### BETWEEN-MAINTENANCE PROCEDURE COMPARISONS

For convenience of exposition, the studies considered in this section are grouped into three categories—those that assess the effect of different food maintenance regimens with a consumatory behavior index, an activity index, or via a measure of instrumental learning.

#### *Consumatory Index*

Marx (1952) investigated the effects of differential infantile deprivation on eating behavior in adult rats. Three groups of 28-day-old subjects were employed; Group C constituted the control (ad-libitum) condition, Group DL was given 10 minutes of free feeding (wet mash) at 8:00 A.M. and 8:00 P.M., and Group DU was given an unlimited amount of time to consume the same amount of food eaten by a matched DL animal. Respectively, these feeding procedures conform to RFTal (Group C), RFT10m/12h (Group DL), and RAA/DL (Group DU) conditions. After 10 consecutive days on these feeding regimens, all subjects were placed on a RFTal pattern for 30 weeks. At the end of this time, the average body weight of all groups was equivalent and adult testing commenced. Testing consisted of giving the subject a single 10-minute free-feeding period once a day (at 1:00 P.M.) for 7 days, followed by a series of hoarding (food pellet) test trials. While no differences in hoarding between the experimental groups were noted, it was found that RFT10m/12h subjects consumed more food than RFTal subjects on the first 2 days of a series of eating-rate test trials. A reliable RAA/DL-RFTal difference was not found. Thus, under the conditions of this experiment, some evidence was found that early food privation experience on a RFT feeding regimen did produce a residual effect on consumatory behavior.

Wilson, Bernstein, and Turrell (1968) attempted to determine the relationship between prolonged exposure to various feeding schedules (when subjects were between 23 and 100 days of age) and alcohol consumption during adulthood. Between 100–124 days of age, the subjects in all groups were placed on a RFTal maintenance schedule, with alcohol consumption testing occurring between 110–124 days, inclusive. Data analysis, with body weight adjusted for all conditions, indicated a number of intergroup differences in alcohol consumption. Specifically, Group I (control condition—RFTal) subjects consumed significantly more alcohol than either Group II (RAA80%—these subjects had been fed 80% of a control subject's daily food intake), or Group III subjects (the latter rats had been administered repeated cycles of 5 days RFTal, 2 days RFTo). No reliable differences between experimental conditions were found.

To summarize briefly, in both the Marx (1952) and Wilson et al. (1968) studies, exposure during infancy to RFT and RAA feeding procedures produced residual effects on consumatory behavior in adulthood when comparison with a RFTal group was made, with weight loss controlled. However, no significant differences between experimental RFT and RAA conditions were found.

#### *Activity Index*

A quite extensive investigation of maintenance schedule effects was conducted by Moskowitz (1959), in which running wheel activity was the behavioral index used to estimate motivational level. In Experiment I, independent groups of Wistar rats were placed on three separate feeding regimens: Each Group A subject was fed 40% of its usual wet mash intake (RAA40%), Group B animals were fed according to a RFT1h/24h schedule, and Group C rats were maintained on a RFP80% regimen. None of these methods was found to be suitable insofar as adaptation to the feeding regimen was concerned (as indexed by body weight change) in conjunction with number of activity wheel revolutions. Thus, Experiment II was conducted to further explore the relationship between activity and the subject's percentage body weight.

In Experiment II, Groups A, B, C, and D were run in a two-phase experiment. The procedures that corresponded to the two phases for the respective groups were RFTal-RFTal (Group A), RFP100%-RFP90% (Group B), RAP80%-RAP90% (Group C), and RFP80%-RFP90% (Group D). In each case, the first feeding-method designation refers to the maintenance schedule in force during Phase 1, while the second designation indicates the Phase 2 procedure. Each subject was allowed one hour per day of unrestricted wheel activity. Results showed that the subject's activity level was most accurately predictable when the RAP procedure (Group C) was employed. This conclusion was buttressed in Experiment III, in which two conditions were employed; a control group (RFTal), and an experimental group (RAP100%-60%; each subject's weight was decreased gradually from 100% to 60% of its weight, adjusted for normal growth). Further, Experiments II and III showed that 85%–90% of the subject's normal weight is the "breaking point" or threshold value at which activity is affected. The conclusion to be drawn from Moskowitz's study is clear: The RAP deprivation procedure is a more appropriate one than the RFT procedure. It can be further added that this conclusion is particularly true when young rats are used since the function relating weight (in grams) and age (in weeks) is a negatively accelerated increasing one (Kaplan & Kaplan, 1965), with the greatest rate of body weight change occurring when the rat is 0–16 weeks of age (Zucker & Zucker, 1942). Thus, the actual weight deficit differences between RAP80% and RFP80% rats 8–12 weeks of age is, for example, much greater than for rats 23–27 weeks of age.

A rather convincing study comparing the effects of RFP and RAA procedures was conducted by Treichler and Collins (1965). Three groups of rats were continuously deprived for 24, 72, or 168 hours until they weighed, respectively, 89%, 80%, and 69% of their predeprivation weights (RFP89%, RFP80%, and RFP69% conditions). Three other groups of rats were placed on a 23-hour deprivation schedule for 20 days and fed adjusted amounts of food in order to bring their percent body weights into line with the RFP subjects. The adjusted groups can be designated RAA89%, RAA80%, and RAA69%. Thus, weight loss was equivalent between RFP and RAA matched pairs. Analysis of running wheel scores showed that at no body weight percentage level did a significant RAA-RFP difference obtain. Thus, as long as current weight loss was equivalent, the manner in which it was lost (as defined by the deprivation procedure employed) was of no consequence.

As reported in an unpublished dissertation, Mathews (1967) examined the differences exerted by RFT, RAA, and IAA procedures on wheel-running activity. The subjects in each group were deprived of food so as to reduce their body weights to 80% of their normal levels (i.e., adjusted for growth). The experimental manipulation consisted of the way this weight reduction was accomplished. The CD subjects were continuously deprived until they weighed the appropriate amount (RFTo procedure), fixed interval (FI) 12 and FI-24 subjects were fed, respectively, every 12 and 24 hours (RAA12h and RAA24h procedures) and variable interval (VI) 12 and VI-24 subjects were fed appropriate amounts of food once on the average of every 12 or 24 hours (IAA12h and IAA24h procedures, respectively). The pattern of obtained results was a rather complex one. The RAA12h and IAA12h subjects made more wheel revolutions than the RFTo subjects; however, the RAA24h and IAA24h subjects did not. It was also found that RAA24h and IAA24h subjects showed a pattern of wheel running that was clearly related to the probability of food presentation as defined by each group's deprivation schedule. Thus, under some "adjusted amount" circumstances, an activity index of hunger drive is modified by the very procedure that induces the drive state.

*Instrumental Learning Index*

Whereas the studies involving consumatory and activity indexes of deprivational experience have been quite similar insofar as the particular task involved is concerned, for example, rate of eating, amount eaten (consumption indexes), wheel running (activity index), the studies in the present category have employed a wide variety of learning situations in which investigation has been made of the effects of maintenance procedures.

In two well-controlled experiments reported by Mandler (1957), the influence of RFA and IFA feeding methods was examined in a single-alternation multiple T-maze problem. In Experiment 1, two groups of rats, Group R (RFA procedure) and Group I (IFA procedure), were formed and placed on different maintenance regimens at 70 days of age. While the subjects of both groups ingested the same amount of food per day, the distinction between them has to do with their patterns of feeding. The RFA subjects were fed every 12 hours, while the IFA subjects experienced deprivation intervals that varied from 12 to 48 hours. It is important to note that the weights of RFA and IFA subjects were either equivalent or the subjects in the latter condition were heavier than subjects in the former. The same two maintenance procedures were employed in Experiment 2; however, the range of IFA deprivation intervals extended from 3 to 30 hours and also the number of IFA feedings equaled the number of RFA feedings. Again, the body weights of the two groups were comparable. The results of both experiments showed that IFA subjects behaved like subjects operating under a higher level of drive than RFA subjects, although this clearly would not have been predicted from the standpoint of the body weight data of the two groups. Specifically, in comparing the current results with earlier data gathered in a similar task, RFA subjects behaved like subjects that had been deprived of food for 12 hours, while IFA subjects behaved like 36-hour deprivates. Thus, RFA and IFA feeding methods did not produce equivalent results in a maze-learning task.

In some follow-up work on this same problem, Mandler (1958) examined the residual, long-term effectiveness of RFT and IFT procedures. In Experiment I, and beginning at 24 days of age, a control group was maintained on a RFT12h/24h schedule. At the same starting age, the rats in the experimental group were placed on an IFT12h/24h feeding regimen, as defined by deprivation intervals that ranged from 6 to 60 hours and free-feeding periods that ranged from 12 to 42 hours in length. From 50 to 85 days of age, all subjects were placed on the RFT12h/24h feeding schedule. During the test period that followed Day 85, all subjects were placed on a standard RFT24h/48h regimen. Thus, the only difference during testing between the control (RFT12h/24h) and experimental (IFT12h/24h) conditions was that the latter had experienced an irregular feeding pattern during infancy while subjects in the former condition had not. Measures of exploratory, consumatory, and instrumental (bar-press) behavior were taken, although the results of the first two measures are difficult to interpret since intergroup body weight differences existed during these test trials. Bar-press results, however, showed that IFT subjects responded at a higher rate than RFT rats on fixed ratio (FR) 10 and FR-20 reinforcement schedules. Importantly, the latter result was achieved when there was no intergroup difference in body weight.

Mandler (1958) conducted Experiment II in order to examine further the relationship between feeding experience in infancy and adult behavior. Between weaning and 58 days of age, control subjects were maintained on a RFTal feeding schedule while the experimental rats were placed on an irregular feeding schedule that averaged approximately 15 hours of free feeding out of every 24 (IFT15h/24h). Between 58 and 106 days of age, all subjects (RFT and IFT) were maintained under RFTal feeding conditions. Subsequent tests of consumatory and bar-press behavior were made when all subjects were under common RFT deprivation conditions, and when the mean body weights of the groups were the same. Results indicated that although there were no differences in exploratory behavior, it was clearly shown that IFT subjects excelled RFT subjects on both consumatory and instrumental behavior indexes. It must be stressed that when these behavioral differences occurred, both groups' current deprivation schedules and weights were equal, the only difference between the two groups being the different feeding regimens established during infancy. Thus, Mandler has shown that, with weight loss held constant, there is a significant RFA-IFA (Mandler, 1957) and RFT-IFT (Mandler, 1958) deprivation schedule effect.

Davenport and Goulet (1964) conducted an investigation that was similar in rationale to that expressed by Ramond et al. (1955). Following habituation to the laboratory, subjects were assigned to one of three experimental conditions: Group I was placed on a RFT30m/24h feeding schedule, Group II was maintained on a RFP80% regimen, and Group III was placed on a RAP80% regimen. After 6 days of habituation to its maintenance schedule, the subject was shaped up to make a bar-press response, and was gradually introduced to, and tested on, a VI2 reinforcement schedule (Testing-A phase). Following Testing-A, subjects in Group II were shifted to the RAP80% maintenance schedule and subjects in Group III to a RFP80% regimen. Groups II and III were each allotted 10 days to habituate to their new schedules. Group I was continued on its original RFT30m/24h feeding regimen. All subjects were then tested on the VI2 schedule of reinforcement for three 30-minute sessions (Testing-B phase). Results showed that body weight data and instrumental response data were highly correlated. Specifically, when either Group II or Group III was on the RAP80% schedule, the subjects' mean body weights were always at 80%, ± 3%, of the weights of equally old nondeprived control subjects, and their bar-press rate as a function of testing days took the form of a flat gradient. Conversely, the subjects under the experimental RFT and RFP conditions (Groups II and III, respectively), although their absolute body weights remained at a rather constant level, continued to drop percentagewise when compared to the weight of ad-libitum control subjects. As might be expected, the bar-press rates of Groups II and III rose over testing days. It should be pointed out that the particular RFT and RFP values (30m/24h and 80%, respectively) did produce equivalent body weights in subjects under these different regimens and likewise produced equivalent bar-press behavior. Thus, equal current weight loss, independent of whether it was produced by the RFT or RFP procedure, was positively correlated with bar-press behavior. The major conclusion of the Davenport and Goulet study was that unless allowance is made for normal growth, a serious misinterpretation can ensue when instrumental behavior is used to index the subject's motivational level. This conclusion is similar to that offered by Moskowitz (1959).

On balance, then, a rather complex picture emerges from studies that have employed instrumental behavior indexes to assess the effect(s) of different maintenance regimens. In a maze-learning task, with weight loss controlled, IFA subjects behaved quite differently from RFA-maintained subjects under different current deprivation conditions (Mandler, 1957). Also, Mandler (1958) showed that rats, under equivalent current deprivation conditions (RFT procedure), but with different deprivation histories, behaved quite differently in terms of consumatory and instrumental behavior indexes. Finally, no RFT-RAP maintenance procedure difference was reflected in bar-press activity, with weight loss held constant (Davenport & Goulet, 1964).

## SUMMARY AND CONCLUSIONS

The primary purpose of this paper has been to examine the possibility that the food maintenance schedule variable may be an important factor in the determination of behavior in a test situation. Since a number of writers (Bolles, 1965; Bolles, 1967; Campbell et al., 1961; Williams & Campbell, 1961) have affirmed that weight loss is the crucial factor that determines hunger drive, an evaluation of the above two positions must involve an examination of the results of studies in which the weight loss and maintenance schedule variables are not confounded. Seven of the studies reviewed herein (Davenport & Goulet, 1964; Mandler, 1957, 1958; Marx, 1952; Mathews, 1967; Treichler & Collins, 1965; Wilson et al., 1968) meet this criterion in that, within a given study, the various groups employed differed from each other only with respect to feeding regimen, weight loss being held constant. In five of the seven studies just referred to (Davenport & Goulet's and Treichler & Collins' being the exceptions), reliable differences between maintenance regimens were reported. Specifically, Mandler (1958), Marx (1952), and Wilson et al. (1968) tested subjects under similar current deprivation conditions, although the pretest maintenance schedules were different for the various groups. They each reported that even with equivalent current weight loss, differences existed as a function of the subject's deprivation history. By contrast, Mandler (1957) and Mathews (1967) assigned rats to different maintenance conditions and kept each subject on the same schedule both before and during testing. Each investigator reported differences between groups that appear to be attributable to the feeding method variable only.

On balance, then, it appears as if the assumption that different feeding methods are equivalent is not tenable. Further, it appears as if weight loss, while probably the most crucial factor, is not the only determinant of the level of hunger drive. It might be that certain maintenance methods are functionally equivalent to one another; however, only future experimentation, in which different maintenance schedules are compared within the same investigation, and in which weight loss is held constant, will permit accurate statements to be made about such equivalence. Until such data are available, it seems appropriate to recommend the use of an adjusted percentage maintenance method in experiments where hunger drive is either manipulated or held at a static level. This recommendation is based principally on the results of Moskowitz (1959), Davenport and Goulet (1964) (these experimenters support the use of the RAP method), and on the work of Ehrenfreund (1966) and Ehrenfreund and Allen (1966), who have studied the behavioral effects of hunger by using the Ehrenfreund weight control device (Ehrenfreund, 1960). As stated earlier, the Ehrenfreund weight control technique entails maintaining the subject on an IFP basis. However, Ehrenfreund (1959, Figure 1) has shown that even mature rats will gain weight consistently if allowed to feed ad libitum. Such being the case, rats should be maintained on an IAP basis in his device so that the subject can always be maintained at a proportion of its "true" percentage body weight level (i.e., adjusted for growth). The appropriateness of both the RAP and IAP methods is due to the fact that each of these schedules allows the experimenter to equate all subjects for percentage weight loss over the entire course of an experiment, irrespective of its length. This eliminates weight loss as a source of confounding. The most common feeding methods currently employed in rat investigations, namely, RFA and RFT, do not control for differential body weight loss, on either an absolute or percentage basis. This, as has been pointed out on a number of occasions earlier in this paper, complicates the interpretation that can be made of experimental findings.

## REFERENCES

Adlerstein, A., & Fehrer, E. The effect of food deprivation on exploratory behavior in a complex maze. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 250–253.

Baker, R. A. The effects of repeated deprivation experience on feeding behavior. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 37–42.

Bare, J. K. Hunger, deprivation, and the day-night cycle. *Journal of Comparative and Physiological Psychology*, 1959, **52**, 129–131.

Bare, J. K., & Cicala, G. Deprivation and time of testing as determinants of food intake. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 151–154.

Beach, F. A. The snark was a boojum. *American Psychologist*, 1950, **5**, 115–124.

Bitterman, M. E. Phyletic differences in learning. *American Psychologist*, 1965, **20**, 396–410.

Bolles, R. C. Consummatory behavior in rats maintained a-periodically. *Journal of Comparative and Physiological Psychology*, 1965, **60**, 239–243.

Bolles, R. C. *Theory of motivation*. New York: Harper & Row, 1967.

Bolles, R. C., & de Lorge, J. Exploration in a Dashiell maze as a function of prior deprivation, current deprivation, and sex. *Canadian Journal of Psychology*, 1962, **16**, 221–227. (a)

Bolles, R. C., & de Lorge, J. The rat's adjustment to a-diurnal feeding cycles. *Journal of Compara-*

*tive and Physiological Psychology,* 1962, **55**, 760–762. (b)

Campbell, B. A., Teghtsoonian, R., & Williams, R. A. Activity, weight loss, and survival time of food-deprived rats as a function of age. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 216–219.

Capaldi, E. J., & Robinson, D. E. Performance and consummatory behavior in the runway and maze as a function of cyclic deprivation. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 159–164.

Davenport, D. G., & Goulet, L. R. Motivational artifact in standard food maintenance schedules. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 237–240.

de Lorge, J., & Bolles, R. C. Effects of food deprivation on exploratory behavior in a novel situation. *Psychological Reports,* 1961, **9**, 599–606.

Duda, J. J., & Bolles, R. C. Effects of prior deprivation, current deprivation, and weight loss on the activity of the hungry rat. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 569–571.

Ehrenfreund, D. The relationship between weight loss during deprivation and food consumption. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 123–125.

Ehrenfreund, D. The motivational effect of a continuous weight loss schedule. *Psychological Reports,* 1960, **6**, 339–345.

Ehrenfreund, D. Weight loss time as a function of constantly maintained percentage body weight. *Psychological Reports,* 1966, **19**, 829–830.

Ehrenfreund, D., & Allen, J. D. Speed scores obtained at random versus standard times under a continuous percentage weight loss. *Journal of Experimental Psychology,* 1966, **72**, 635–639.

Ehrenfreund, D., & Badia, P. Response strength as a function of drive level and pre- and postshift incentive magnitude. *Journal of Experimental Psychology,* 1962, **63**, 468–471.

Eisman, E. An investigation of the parameters defining drive (D). *Journal of Experimental Psychology,* 1956, **52**, 85–89.

Eisman, E., Asimow, A., & Maltzman, I. Habit strength as a function of drive in a brightness discrimination problem. *Journal of Experimental Psychology,* 1956, **52**, 58–64.

Eisman, E., Linton, M., & Theios, J. The relationship between response strength and one parameter of the hunger drive. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 359–363.

Eisman, E., Theios, J., & Linton, M. Habit strength as a function of drive in a bar-pressing situation. *Psychological Reports,* 1961, **9**, 583–590.

Fallon, D. Effects of cyclic deprivation upon consummatory behavior: The role of deprivation history. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 283–287.

Hoyenga, K. T., & Aeschleman, S. Social facilitation of eating in the rat. *Psychonomic Science,* 1969, **14**, 239, 241.

Hughes, R. N. Effects of food deprivation, deprivation experience and sex on exploration in rats. *British Journal of Psychology,* 1968, **59**, 47–53.

Jensen, G. D. Learning and performance as functions of ration size, hours of privation, and effort requirement. *Journal of Experimental Psychology,* 1960, **59**, 261–268.

Kaplan, M., & Kaplan, E. F. Estimating growth of rats for behavioral and biological study. *Psychological Reports,* 1965, **17**, 419–423.

Lawrence, D. H., & Mason, W. A. Food intake in the rat as a function of deprivation intervals and feeding rhythms. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 267–271.

Mandler, J. M. Irregular maintenance schedule and drive. *Science,* 1957, **126**, 505–507.

Mandler, J. M. Effect of early food deprivation on adult behavior in the rat. *Journal of Comparative and Physiological Psychology,* 1958, **51**, 513–517.

Marx, M. H. Infantile deprivation and adult behavior in the rat: Retention of increased rate of eating. *Journal of Comparative and Physiological Psychology,* 1952, **45**, 43–49.

Mathews, S. R., Jr. *The effect of schedule of deprivation on wheelrunning.* (Doctoral dissertation, University of Virginia) Ann Arbor, Mich.: University Microfilms, 1967, No. 67-17, 612.

Mirsky, A. F., Looney, E., & Rosvold, H. E. Maze retention deficit following electroconvulsive shock independent of feeding schedule. *Journal of Comparative and Physiological Psychology,* 1954, **48**, 403–405.

Moll, R. P. Drive and maturation effects in the development of consummatory behavior. *Psychological Reports,* 1964, **15**, 295–302.

Moskowitz, M. J. Running-wheel activity in the white rat as a function of combined food and water deprivation. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 621–625.

Olewine, D. A., Barrows, C. H., Jr., & Shock, N. W. Effect of reduced dietary intake on random and voluntary activity in male rats. *Journal of Gerontology,* 1964, **19**, 230–233.

Perin, C. T. Behavioral potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *Journal of Experimental Psychology,* 1942, **30**, 93–113.

Ramond, C. K., Carlton, P. L., & McAllister, W. R. Feeding method, body weight, and performance in instrumental learning. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 294–298.

Siegel, P. S. Food intake in the rat in relation to the dark-light cycle. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 294–301.

Stolurow, L. M. Rodent behavior in the presence of barriers: II. The metabolic maintenance method; a technique for caloric drive control and manipulation. *Journal of Genetic Psychology,* 1951, **79**, 289–335.

Strong, P. N., Jr. Activity in the white rat as a function of apparatus and hunger. *Journal of*

*Comparative and Physiological Psychology*, 1957, **50**, 596–600.

TREICHLER, F. R., & COLLINS, R. W. Comparison of cyclic and continuous deprivation effects on wheel running. *Journal of Comparative and Physiological Psychology*, 1965, **60**, 447–448.

TREICHLER, F. R., & HALL, J. F. The relationship between deprivation weight loss and several measures of activity. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 346–349.

VERPLANCK, W. S., & HAYES, J. R. Eating and drinking as a function of maintenance schedule. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 327–333.

WARNER, L. H. A study of hunger behavior in the white rat by means of the obstruction method. *Journal of Comparative Psychology*, 1928, **8**, 273–299.

WEASNER, M. H., FINGER, F. W., & REID, L. S. Activity changes under food deprivation as a function of recording device. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 470–474.

WILLIAMS, R. A., & CAMPBELL, B. A. Weight loss and quinine-milk ingestion as measures of "hunger" in infant and adult rats. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 220–222.

WILSON, A. S., BERNSTEIN, L., & TURRELL, E. S. Alcohol selection in rats with different early feeding experiences. *Quarterly Journal of Studies on Alcohol*, 1968, **29**, 566–570.

WOODS, P. J., & BOLLES, R. C. Effects of current hunger and prior eating habits on exploratory behavior. *Journal of Comparative and Physiological Psychology*, 1965, **59**, 141–143.

YOUNG, P. T. *Motivation and emotion: A survey of the determinants of human and animal activity.* New York: Wiley, 1961.

ZIMBARDO, P. G., & MILLER, N. E. Facilitation of exploration by hunger in rats. *Journal of Comparative and Physiological Psychology*, 1953, **51**, 43–46.

ZIMBARDO, P. G., & MONTGOMERY, K. C. The relative strengths of consummatory responses in hunger, thirst and exploratory drive. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 504–508.

ZUCKER, L., & ZUCKER, T. F. A simple time weight relation observed in well nourished rats. *Journal of General Physiology*, 1942, **25**, 445–463.

(Received March 10, 1971)

---

## Manuscripts Accepted for Publication in the

## *Psychological Bulletin*

Implicit Personality Theory: A Review. David J. Schneider (Department of Psychology, Brandeis University, Waltham, Massachusetts 02154).

VOL. 78, NO. 5 NOVEMBER 1972

Psychological Bulletin

Diary No. 734(3)
Date 4.2
il. N: 216
Bureau Ednl Psy. Research.

# MALNUTRITION AND MENTAL DEFICIENCY[1]

BONNIE J. KAPLAN[2]

*Brandeis University*

Inadequate nutrition can affect both physiological and psychological development in humans. The nine months of gestation and the first few years of life are the most critical in the growth of brain tissue, and are also the periods of greatest vulnerability to malnutrition. Some variables that mediate the effect of nutrition on mental development include prematurity, birth weight, and the nutritional status of the mother during her childhood. A model for thorough research is suggested and is used to evaluate the major studies of the relationship between malnutrition and mental deficiency.

This review examines the interaction of malnutrition and mental deficiency. Although the evidence on nonhuman species shows fairly conclusively that malnutrition retards mental development, no definitive study on humans has yet been done. The ethical and methodological difficulties involved in studying the malnutrition of humans are several: We cannot intentionally subject an experimental group of children to a low-protein diet and thereby risk the possibility of retarding the child for life; we cannot be certain to what extent a child has been malnourished in his home environment since diets can fluctuate daily; and we cannot even be certain that a relatively undernourished and mentally subnormal child is, in fact, subnormal because of his malnutrition.

Although a causal connection has not been directly demonstrated, the research reviewed contains compelling evidence that early malnutrition is a contributing factor in the incidence of mental deficiency. First, malnutrition is discussed as a physiological entity, affecting physical health and the development of the central nervous system. Next, the evidence relating malnutrition and mental deficiency is analyzed, comparing existing research to six criteria that should be fulfilled in an adequate research design. Finally, the problem is discussed in relation to social factors and relevance for psychologists.

## MALNUTRITION AS A PHYSIOLOGICAL ENTITY

### *Specific Vitamin and Mineral Deficiencies*

A great deal is known about the role of specific nutrients in affecting physical development. Beriberi is the disease that results from an insufficient supply of thiamine; pellagra results from insufficient nicotinic acid; rickets results from insufficient vitamin D; and scurvy results from a deficient supply of vitamin C (Crome & Stern, 1967). In terms of psychological effects, insufficient iodine in a pregnant woman's diet can cause her to give birth to a child suffering from endemic cretinism, which involves extreme mental deficiency (Robinson & Robinson, 1965).

It is not surprising to find that nutritional anemias, such as those caused by a lack of folic acid or vitamin $B_{12}$, are significantly more common in low-income groups than in other classes (Klipstein, 1968). This unequal distribution of anemias is obviously due to the unequal distribution of nutritious food. This type of deficiency is one example of a common phenomenon: Deficiencies in a woman's diet may have a minimal observable effect on her own physical health, but the deficiency

---

[1] Preparation of this review was supported in part by Public Health Service Grant 2 F01 MH 39,428-02 MTLH.

[2] Requests for reprints should be sent to Bonnie J. Kaplan, Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.

© 1972 by the American Psychological Association, Inc.

can severely affect her offspring. Some anemias, such as iron deficiency anemia in infants, can be the direct result of iron deficiency in the mother's diet, specifically during her pregnancy (Strauss, 1953). There is also evidence, however, that iron deficiency in the pregnant woman's diet can have a much more deleterious effect than the production of anemia. McGanity, Bridgforth, and Darby (1958) found a significantly higher rate of serious congenital abnormalities in babies of mothers suffering from iron deficiency anemia than in nonanemic mothers.

### *Marasmus and Kwashiorkor: Two Types of Protein-Calorie Malnutrition*

Marasmus is a disease resulting from a diet insufficient in both calories and proteins. It usually occurs in children between the ages of 6 and 18 months (Gopalan, 1967). Its two major characteristics are severe weight loss in all body organs and subcutaneous fat, plus mental apathy (Chatterjee & Mukherjee, 1968).

Kwashiorkor has been interpreted both as a disease entirely separate from marasmus and as a more severe form of marasmus that occurs later in life. Kwashiorkor results from "prolonged subsistence on diets which [are] . . . moderately poor in calories but very poor in protein [Chatterjee & Mukherjee, 1968, p. 145]." It usually occurs between the ages of 12 and 36 months, though it can appear later. Although the severe wasting found with marasmus is not apparent, there is some loss of weight, especially involving muscle tissue. Kwashiorkor's major characteristics are severe skin disorders, excessive infiltration of fat into the liver (thus enlarging the liver and causing the commonly seen protruberant belly), low plasma proteins (and consequent edema, which is the subcutaneous fluid often noted), mental apathy, and the discoloration and loss of hair.

In both these diseases the child's condition is usually worsened by infectious diseases. The death rates of children aged 1–4 years in countries such as Mexico, Guatemala, and Peru are often 40 times the rate for the same age group in the United States. This difference reflects mainly the problems of malnutrition and subsequent vulnerability to infectious disease and parasites (Jelliffe, 1968). In cases where children are inadequately spaced, the difficulties are exacerbated. Each child must be separated from the breast, usually the most reliable source of nutrition for the newborn, in order to make room for the next child. In fact, the name "kwashiorkor" actually comes from a word in the Ga language of West Africa that means "disease that occurs when displaced from the breast by another child [Jelliffe, 1968, p. 76]."

Mental apathy is characteristic of both marasmus and kwashiorkor. The return of interest in the environment and responsivity to external stimuli are often used as indexes of the effectiveness of treatment for the diseases.

### *Physiological Changes Attributed to Nutritional Deficiencies*

Some of the most interesting, although necessarily less rigorous, studies of undernutrition are those that examine the effects of wars and depression. Montagu (1962) summarized 29 studies of malnutrition during World Wars I and II, all of which showed that the average birth weights of infants born during and immediately after the war periods were significantly lower than during nonwar years.

Antonov's (1947) study of the German siege of Leningrad was one of the most complete reports. The siege lasted from August 1941 to January 1943. The food supply was extremely low for most of that time, and there was no heat for most homes, even during the coldest winter months. There were also the added stresses of daily air raids and shelling. During the worst part of the siege, the prematurity rate was 41.2%; the mortality rate of full-term births was 9% and of premature births was 30.8%. All of these figures are extraordinarily high. Antonov, minimizing the role of stress, attributed these problems—high rate of premature births, high mortality rates of full-term and premature births—to intrauterine undernutrition of the fetus.

Laporte (1946) studied the heights and weights of Paris grade school children in 1944 and compared the data to those of the same age groups in 1938. He found there was a significant decrease corresponding to the war years. Again, the results were attributed to inadequate nutrition.

Palmer (1935) did a longitudinal analysis of three groups of Americans who experienced the depression: children who were not poor and whose families did not suffer from the depression, children who were poor and who remained poor, and children whose families experienced extreme economic losses due to the depression. He found that the first group remained above average in height and weight; the second group remained below average; but the third group dropped from above average to below average. Once again, the author attributed the changes to decreased nutrition.

The fact that Antonov, Laporte, and Palmer attributed their results to decreased nutrition does not rule out the possibility that stress was the crucial factor. Montagu (1962), in reviewing a series of German studies done following World War II, concluded that the general wartime decrease in birth weights and increase in congenital malformations were due to malnutrition and stress. That stress does increase during wars and depression is obvious. Perhaps the results of these studies cannot be interpreted entirely in terms of nutrition.

Since there is no direct evidence to determine which variable is more important—malnutrition or stress—we might turn to a different type of research for an indirect answer. There are many studies in the literature that compare different socioeconomic groups (Anderson, Baird, & Thomson, 1958; McCance, Widdowson, & Verdon-Roe, 1938; Montagu, 1962; Walker, 1955). In virtually all such studies the following relationship seems to hold: Relatively poorer women, when compared to women of higher economic classes, have less nutritious diets and give birth to babies who are more often of low birth weight or perhaps premature, are less healthy, are more likely to have congenital malformations or serious diseases, and have a somewhat lower probability of living beyond a few months. It can be argued that women of lower classes are also under more stress than women of higher economic classes, but a series of experiments has proven that dietary supplements given to pregnant women of low socioeconomic groups have a significant effect in improving the health of their babies (Dieckmann, Turner, Meiller, Savage, Hill, Straube, Pottinger, & Rynkiewicz, 1951; Ebbs, Brown, Tisdall, Moyle, & Bell, 1942; Toverud, Stearns, & Macy, 1950). In these experiments the economic and social problems presumably creating the stresses in their lives were not treated. Ebbs et al. (1942), in a Toronto prenatal clinic, studied 120 pregnant women on low protein-calorie diets (their normal diets) and 90 women on the same diet who were given supplements. All the women were of low economic status. The authors concluded that on every physical measure the babies of mothers given the supplements were superior to the babies of the other group.

This line of argument leads one to the conclusion that the role of stress within the intrauterine environment is not known, but that intrauterine nutrition seems to be very important.

The relationship between social class and nutrition seems clear. Now what of the relationship between nutrition and problems of pregnancy and birth? Toverud et al. (1950) reviewed a large number of studies in which vitamin and mineral intakes of pregnant women were controlled. The overall results were that the control of vitamins and minerals led to significant decreases of stillbirth and neonatal mortality rates. McCance et al. (1938) compared pregnant wives of unemployed English miners to pregnant wives of professionals. They concluded that the intake of vitamins and minerals was related to the economic income and social class of the women. Consumption of calories, fats, and carbohydrates was not at all related to income, but protein consumption was definitely associated with economic status. This finding is important because Burke, Harding, and Stuart (1943) then found a consistent parallel between protein intake of mothers during the last 6 months of pregnancy and the heights and weights of the infants. Dieckmann et al. (1951) also noted a significant increase in the physical health of babies as the consumption of protein was increased during the mother's pregnancy.

The Harvard School of Public Health along with two Boston hospitals conducted a longitudinal study of 200 women and their pregnancies (Kirkwood, 1955). After following the women for 20 years, the researchers concluded that a good prenatal diet produced the best chance of insuring the good physical condition of the child. Nutrition of the mother was considered important not only during pregnancy but also when she herself was a child. As Kirkwood summarized: "Good obstetrical care [including good diet] is a very broad type of care that must start with the birth of the future mother, or even before that, with her mother who is the grandmother of the baby we are discussing [1955, p. 31]."

Birch[3] has discussed the intergenerational aspects of this problem. Citing data from a series of studies in Aberdeen, Scotland (Baird, Thomson, & Duncan, 1953; Baird, Walker, & Thomson, 1954), Birch claimed that low birth weights of infants were highly correlated with mothers' heights, which were in turn correlated with the mothers' socioeconomic status when the mothers were born, not with the mothers' socioeconomic status at the time of pregnancy. (In Aberdeen all women receive comparable health care because of the national health care system, and complete records are kept on all individuals throughout their lives. These records permitted accurate assessment of social class.)

One mechanism by which the intergenerational influences of socioeconomic status might affect pregnancy and childbirth was suggested by Walker (1955). He pointed out that in the Aberdeen studies perinatal mortality (which includes stillbirths and babies who die in the first week after birth) was more frequent in the lower social classes, and even more frequent in small women. He attributed the role of stature to the structure of the pelvic region:

> Good environment and adequate nutrition during childhood should prevent flattening of the bony pelvis. Full nutrition during prepuberty and puberty phases should allow the child to attain full potential height and physique, and the pelvis its maximal capacity [p. 30].

Thomson (1959) studied what he termed the "reproductive efficiency" of women. He found that there was an inverse relationship between the mother's stature (and presumably pelvic size) and problems of pregnancy such as fetal malformations, hemorrhage, and perinatal deaths.

Jeans, Smith, and Stearns (1955) collected data on 400 pregnant, low-income women in Iowa, whose diets were commonly deficient in calcium and protein. Besides the relationship between mothers' nutrition and birth weight of babies, as also reported by Burke et al. (1943), there were significant relationships between poor diet and prematurity rates (more than twice as high as the prematurity rate of the control group with good diets) and also between poor diet and multiple congenital anomalies. Again referring to data from the Aberdeen research, we can conclude that women from the lower social classes (especially those of small stature) tend to have the least nutritious diets and the highest rates of congenital malformations, especially those malformations affecting the central nervous system (anencephaly, hydrocephaly, spina bifida, and so forth) (Montagu, 1962).

The data that have been reported in this subsection suggest more than a relationship between undernutrition and physical deficiencies. They suggest also a mechanism by which undernutrition may cause difficulties in pregnancy and birth (viz., reduced pelvic development), and they also demonstrate that the central nervous system is frequently the part of the body that is damaged.

### *Brain Growth and Myelination*

In the previous subsection evidence was presented that links prematurity rates, perinatal deaths, low birth weights, and congenital malformations to nutritional deficiencies in the mother's diet during pregnancy and even earlier. The congenital anomalies, as pointed out by Jeans et al. (1955), are mainly those that involve the central nervous system. Since the main concern in this review is with the psychological effects of malnutrition, the literature on brain and spinal cord defects is discussed in detail.

---

[3] H. G. Birch. Nutritional risk and learning disabilities: An epidemiological view. Symposium presented at the meeting of the American Association for the Advancement of Science, Boston, December 1969.

A number of studies have been done on dogs (Platt, Heard, & Stewart, 1964), pigs (Dickerson, Dobbing, & McCance, 1967), and rats (Davison & Dobbing, 1966; Winick & Noble, 1966), in which the animals have been fed diets that were deficient in some of the basic nutrients. When the animals were sacrificed and their brain tissue analyzed, in all cases the experimenters reported significantly retarded brain development.

It appears that brain growth may be retarded through a number of different mechanisms. Davison and Dobbing (1966) noted that when the undernutrition occurred during the myelination period of development of the rat's brain, they obtained the most severe growth retardation. These researchers believed that inadequate myelination in humans might be a significant factor in mental deficiency, and they emphasized that the impaired intellectual development of children suffering from phenylketonuria is associated with (though not necessarily caused by) defective myelination in the brain.

Another mechanism by which brain growth is retarded in the rat is through insufficient cell division. Winick and Noble (1966) found that the effects of malnutrition on brain cells depended on the phase of growth of the animal at the time of the dietary insufficiencies. For rats, very early malnutrition obstructed cell division; even when the rats were subsequently fed on nutritious diets, they did not recover fully. In general, Winick and Noble found that the later the period of malnutrition occurred, the more likely it was the cell division was unimpeded; later malnutrition did produce decreased cell size, but with subsequent nutritious diets the animals could recover.

The animal studies clearly support the view that the earlier the malnutrition occurs, the more deleterious is the effect on brain growth. There are very few reports in the literature of attempts to verify this conclusion in humans. Although it is not ethical to induce experimentally severe malnutrition in newborn babies, it is possible to do autopsies on children who die of malnutrition. Winick, Rosso, and Waterlow (1970) have reported on the autopsies of 16 Chilean and Jamaican babies who died of malnutrition before 2 years of age. All the children had a subnormal number of brain cells; three of them had less than 40% of the normal number. According to Winick (1968, 1970), the human brain grows mainly by cell division during gestation and the first 6 months of life; after 6 months of age the brain grows mainly by taking in proteins and fats, without increasing the cell number. He concluded:

> Severe undernutrition in the neonatal period will curtail cell division in all organs including brain. Recovery will occur only if adequate feeding is resumed early, presumably while cell division is still occurring [1968, p. 197].

Verification for Winick's results can be found in several other autopsy studies. Brown's (1966) autopsy report on organ weight in 1,094 Ugandan children, who died while suffering from various degrees of malnutrition, demonstrated that brain weight was low for malnourished children. Naeye, Diener, and Dellinger (1969) chose an urban American sample of 252 babies either stillborn or dead within 48 hours of birth. Autopsy results showed that those from poor families were 15% smaller than those from nonpoor environments. All the body organs and the brain weighed significantly less in the infants from poor families. Parekh, Pherwani, Udani, and Mukherjee (1970) found a 28%–36% weight deficit in the autopsied brains of 357 malnourished Indian children. The greater the nutritional deficiency, the greater was the weight deficit, especially if the malnutrition occurred in the last 3 months of gestation or the first 3 months after birth.

Since the present interest is mainly the psychological effects of nutritional deficiencies, the obvious question arises: What parts of the brain seem to be most traumatized in the autopsied tissue? This issue was discussed briefly following one of the American Association for the Advancement of Science symposia on malnutrition and behavior, December 1969. The evidence seems to be inconclusive. Winick and others have found a general reduction in brain weight, with no particular area more significantly affected than any other.

## RELATIONSHIP BETWEEN MALNUTRITION AND MENTAL DEFICIENCY

Some of the research reported in the previous section demonstrated that malnutrition can affect not only prematurity, perinatal deaths, and birth size but also congenital anomalies of the central nervous system and brain weight. Early malnutrition appears to affect cell number in the brain and is not easily overcome; later malnutrition (after approximately 6 months of age in humans) affects brain cell size and is more easily reversed.

What are the behavioral and mental effects of malnutrition? The question is most easily answered in terms of nonhuman species. Changes in the postures, gaits, brain weights, electroencephalographic records, exploratory activity, and learning ability of rats, pigs, dogs, and other animals have been well documented (Scrimshaw & Gordon, 1968). Typical are the results of Caldwell and Churchill (1967); protein-deficient rats had lower birth weights, had a higher preweaning mortality rate, and performed less well on the Lashley III water maze than well-nourished rats.

Mental apathy and psychological deficiency can be the results of insufficiencies of specific vitamins, proteins, and calories, and also of congenital malformations. Having demonstrated that malnutrition is also related to birth weight and prematurity rates, can we show that birth weight and prematurity rates are related to mental deficiency?

### *Prematurity, Birth Weight, and Mental Development*

Knobloch, Rider, Harper, and Pasamanick (1956) studied a variety of factors related to physical and mental development in premature infants. They defined a "continuum of casualty" which ranged from more severe problems such as spontaneous abortion and stillbirths to less severe results like cerebral palsy, epilepsy, and mental deficiency. Comparing 500 premature births with 492 full-term controls, matched for socioeconomic status, season of birth, and parity of the mother, the authors then administered physical and Gesell developmental examinations at about 40 weeks after birth. In general, they found that the incidence of physical and mental abnormalities increased as birth weight decreased. With regard to intellectual potential, they found that whereas 21.8% of the controls were above average, only 16.3% of the moderately low birth weight babies (1,500–2,500 grams) and only 5.3% of the very low birth weight babies (less than 1,500 grams) were above average. Judged as mentally defective or borderline defective were 1.6% of the controls, 1.8% of the moderately small babies, and 17.6% of the very small group. In sum, the probability of a very small baby being completely normal was only 50.9%, compared to 88.4% for the controls. Knobloch et al. (1956) did not control for several variables that might have influenced their data (e.g., the intellectual potential of parents), but the study as a whole was quite complete and the evidence is dramatically clear: babies of low birth weights have a significantly lower probability of being entirely physically and psychologically healthy than babies of normal birth weights.

Using the Gesell schedule, Cravioto, Licardie, Montiel, and Birch (1967) tested motor and adaptive development of 95 premature Guatemalan babies. They excluded from their sample any baby with physiological complications, yet the tests still showed a significant lag in performance when compared to babies born at full term. There was apparently no follow-up study to determine the permanency of the lag. Churchill (1965) used IQ as a measure of the development of a sample of American babies and found a significant negative correlation in twins between IQ and birth weight. Churchill's sample consisted of full-term, low birth weight babies; his results are consistent with Cravioto's.

Cravioto and DeLicardie (1968) were interested in the role of social variables in the relationship between size and intelligence. They compared a group of rural Guatemalan children with a sample of upper-class urban Guatemalans. Height of the children was highly correlated with intelligence in the former group, but not in the latter. The implication of this research, which has been verified in several countries, is that where malnutrition is not a problem, a child's size is determined mainly by genetic factors. In rural areas where malnutrition is more common, however, children who are malnourished never achieve their full height potential, and this stunting of growth is accompanied by mental deficits. Consequently, smaller children also tend to have lower IQs. This type of research is strong confirmation that good nutrition enables a child to reach his maximum genetic potential.

This relationship between birth weight and intelligence has been examined extensively in twins. Willerman and Churchill (1967) administered intelligence tests to 27 sets of identical twins, aged 5 to 15 years. In all cases the twin member who was the smaller of the pair at birth achieved lower scores in verbal and performance IQ. All the twins had normal, full-term deliveries with no complications. In other words, no twin member would have been defined as premature, low birth weight, or physically "at risk." Yet the differences in IQ were statistically significant, and the results were consistent across all pairs studied.

One interpretation of Willerman and Churchill's data is that unequal prenatal sharing may occur such that one twin receives more nutrients than the other. There is evidence for this theory in the "twin transfusion syndrome" (Driscoll, 1964). Because of the sharing of vasculature common among identical twins, one twin sometimes receives more blood (and, consequently, nutrients) than the other. The smaller twin may suffer to the extent of becoming anemic, while the larger fetus frequently has an abundance of blood.

Babson, Kangas, Young, and Bramhall (1964) confirmed Willerman and Churchill's findings: In twins born with widely disparate birth weights, but with no apparent physiological disorders, the smaller twin member consistently had a lower IQ than the larger twin.

Several research teams have done retrospective studies of mentally subnormal children. Pasamanick and Lilienfeld (1955) collected data on mentally defective children born in Baltimore between 1935 and 1952. They compared them with mentally normal children born during the same period. There were significantly more complications of pregnancy and significantly higher rates of prematurity and neonatal anomalies in the mentally defective children. As previously noted, these latter factors are linked with nutritional deficiencies. Birch (1968) reported similar data on prematurity rates.

Churchill, Neff, and Caldwell (1966) pursued the issue of the relationship between birth weight and intelligence. Comparing 51 normal and 51 moderately retarded children, they concluded that the retarded group had significantly lower birth weights than the normal children. All the children were of comparable socioeconomic status. The results could have been attributed to differential prematurity rates, since there is a high prematurity rate among mentally retarded children. When premature births were excluded from the analysis, however, there was still a significant relationship between mental retardation and low birth weight. Churchill et al. (1966) suggested that their results were probably due to intrauterine nutritional variations.

### *Direct Evidence Linking Malnutrition and Mental Deficiency*

What this author considers "direct evidence" are studies that examine and/or manipulate nutrition and assess mental development or IQ in humans. This group of experiments differs from those that are concerned with related or intervening variables (prematurity rates, birth weight, and so forth).

An adequate model for a research design to answer the question, "What is the effect of malnutrition on mental development?" should strive to fulfill the following criteria:

1. Both the duration of malnutrition and the extent of physiological and psychological effects should be assessed.

2. The approximate period in life during which the malnutrition was experienced should be consistent across subjects.

3. All other physiological variables—prematurity, chronic disease, congenital malformations—should be carefully controlled.

4. Environmental factors—family income, socioeconomic class, diet, parents' education—should be carefully controlled.

5. Mental development should be assessed

with a variety of measures: tests for IQ, for cognitive development, for perceptual and motor skills, and perhaps for emotional and personality development.

6. Tests of psychological development should be given periodically over a decade or more, to examine the permanency of the effects.

All of the studies reviewed in this section fail in some way to meet these criteria. This statement does not justify ignoring the research. Several of the studies—notably Chase and Martin (1970), Cabak and Najdanvic (1965), Pollitt and Granoff (1967), and Stoch and Smythe (1967)—present relatively thorough methodologies and very convincing results.

One of the earliest attempts in the United States was the Harrell, Woodyard, and Gates study reported in 1955 (for discussions, see Masland, 1958; Montagu, 1962; Robinson & Robinson, 1965). Harrell et al. chose two samples of pregnant women and administered four types of dietary supplements. The Kentucky sample consisted of 1,200 white women, most of whom were already on fairly adequate diets. The Virginia sample consisted of 1,200 women, both black and white, of poor economic status. The four daily dietary supplements were as follows: (*a*) ascorbic acid; (*b*) thiamine, riboflavin, niacin, and iron; (*c*) inert material; and (*d*) thiamine. Each sample was divided into four groups to receive the four supplements. In the Virginia sample, 518 of the offspring were tested at age 3; 811 of the Kentucky offspring were tested. There were no significant differences in the Kentucky children's IQs, perhaps due to the generally adequate diets of the mothers even without the supplements. In Virginia, however, the group receiving Supplement *b* had a significantly higher average IQ than the group receiving Supplement *c*.

The conclusions drawn from this research are limited, yet still significant. The dietary control was restricted to the supplements, so we are unsure whether the women's normal diets were exactly comparable. The reasons for testing less than half of the Kentucky children are unclear, and there might be some contaminating factors that contributed to the selection of these 518 out of the 1,200. And as Harrell et al. (1955) admitted, the best conclusion that can be drawn from their research is that vitamin supplements can yield a higher IQ until age 4. Children were not followed beyond that age to test the stability of the results. Thus, this experiment did not control for socioeconomic factors in the Kentucky sample, and they failed to examine the permanency of the effects.

Another study that fails to provide complete control of diet is the Christakis, Miridjanian, Nath, Khurana, Cowell, Archer, Frank, Ziffer, Baker, and James (1968) investigation of 642 New York City fifth- and sixth-grade children from the Lower East Side of Manhattan. Each child's nutritional status was determined from subjective reports, while reading test scores provided the index of mental ability. In the entire sample only 6.4% were judged to have excellent dietary histories. In general there was a positive correlation between relatively high reading scores and excellent diet patterns, with Chinese children at the high end on both variables and Puerto Ricans at the low end. Although there were few clinical signs of malnutrition, the researchers did identify a cluster of Puerto Rican children with "suboptimal" nutrition and very low reading scores. This research does not attempt to fulfill any of the six criteria outlined above. It is an epidemiologic survey that depends entirely on subjective reports of dietary habits and on reading ability as a measure of intellectual capacity.

One American study of severe malnutrition was reported by Chase and Martin (1970). Choosing 19 children who had been hospitalized for malnutrition before the age of 1, they tested the children from 2 to 5 years after treatment. Height, weight, head circumference, and developmental quotients were all significantly low for these children. Chase and Martin fulfilled all the criteria outlined above for physiological and environmental controls. Perhaps in a few years they will be able to report further follow-up data.

Cabak and Najdanvic (1965) were concerned about the depression of IQ that they felt followed in severe cases of infant malnutrition. They chose a sample of 36 children in Sarajevo, Yugoslavia, who had been hospitalized between the ages of 4 months and 24 months for treatment of marasmus. At the time of the study, all the children seemed to be physically normal; those with chronic diseases and other defects were excluded from the sample. About two-thirds of the children were the offspring of unskilled laborers; the remaining one-third represented families of higher economic status. When the children were tested with a modified Binet test, only 18 of them scored in the normal range (91–110); 12 scored between 71 and 90; the remaining 6 scored below 70. In other words, not one had an IQ above average, and exactly one-half of them were mentally deficient. The results were significantly lower than Serbian norms. Unfortunately, Cabak and Najdanvic failed to use a control group matched for environmental variables. Birch and Cravioto (1968) obtained similar results with Guatemalan children who had experienced malnutrition. There was an increased tendency for the children to fall in the borderline or subnormal IQ range.

Pollitt and Granoff (1967) conducted a study very similar to Cabak's, using 27 children from the slum section of Lima, Peru. All the children, aged 11–32 months, had been hospitalized for severe protein-calorie malnutrition. When they had apparently recovered from the malnutrition (judged by physiological measures), they were given the Bayley Infant Scales of Mental and Motor Development. Only two of the children scored above the basal levels of the scales; there was severe mental deficiency apparent despite the physiological recovery. To exclude the effects of social and economic factors, Pollitt and Granoff used siblings of the subjects as controls. In all cases, the siblings scored within the normal range. The retardation of the malnourished children seems to be explained best in terms of dietary deficiencies.

One fault in the preceding studies—Chase and Martin, Cabak and Najdanvic, Pollitt and Granoff—is the lack of long-term follow-up data. Are the impairments irreversible? With subsequent adequate nutrition and education for 10 years or so, could the mental deficiencies of these children disappear? This is a crucial question, and so far it has not been answered satisfactorily. Cravioto and Robles (1965) believe that the possibilities for complete recovery from severe protein-calorie malnutrition are slim if the malnutrition occurs before the child is 6 months old. This estimate coincides well with Winick's (1968) research, which showed that total recovery from early malnutrition can occur only if rehabilitation begins during the first 6 months of life, while cell division is still taking place in the brain. Thus, if malnutrition does occur before 6 months of age, and is not treated during the period of cell division, it seems that the brain will be irreparably damaged. Presumably, later malnutrition interferes with protein synthesis, which is easier to correct.

There is one major study in the literature that reports the only significant attempt so far in the world to collect long-term follow-up data. Since the long-term effects of malnutrition seem to comprise a major gap in our knowledge, this research deserves detailed consideration.

Stoch and Smythe (1967) began their South African study in 1955. Twenty black infants who had suffered from extreme protein-calorie malnutrition were matched for sex and age with controls. A large variety of tests were used for the children: Gesell infant scales, electroencephalographic evaluations, and perceptual tasks. An IQ score was estimated from these tests every 2 years.

As might be expected, the heights and weights of the malnourished group were significantly less than the controls. Head circumference was also significantly less; this is an important finding, because there is generally a very high correlation between head circumference and brain weight. Thus, the experimental children seem to have had much smaller brains than the controls. Stoch and Smythe (1967) estimated that the average intracranial volume of the experimental group was 13.7% less than that of the controls.

The majority of the experimental infants (12 of the 20) had lower IQs than the very lowest child from the control group. All the experimental children scored significantly lower than the average control score.

It is worth noting that the control group was itself not above average. They were from the lowest socioeconomic class (as were the experimental children) and their mean weight at age 1 year was below normal. The living conditions of the two groups in the beginning of the study did differ, however, with the experimental children more commonly living in broken homes, with a higher rate of alcoholism and unemployment. By 1967 most of the experimental group's home environments had improved. The initial differences, however, indicate poor control of environmental factors. Stoch and Smythe considered the environmental effect in one possible explanation of their data. They suggested that the inactivity and lack of energy associated with malnutrition, plus an inadequate home environment, could have combined to decrease the receptiveness of the malnourished children to stimuli and therefore have impaired the learning process. Their alternative hypothesis was that early malnutrition causes brain damage by retarding brain growth.

There are a number of reasons why Stoch and Smythe favor the second explanation. Many of the results from the nonverbal tests (e.g., visual-motor and pattern perception) were very similar to the tests of children suffering from brain damage. The abnormal electroencephalograms and the decreased head circumferences also led them to favor the organic damage interpretation. And finally, the children had experienced undernutrition during the period of maximum brain growth and development, namely, before the age of 2 years. They concluded that the first 2 years of life are crucial.

These 2 years are the years of maximum brain growth; they are also the years during which malnutrition occurs most frequently in all countries of the world. According to Thomson (1968), the situation is worsening in many countries because of an increasing tendency to wean children from the breast at an earlier age. Breast feeding is entirely adequate for children for many months, especially in underdeveloped nations where there is often no other adequately nutritious food available. If the child is deprived of the breast after only 3 or 4 months, and if no other adequate food source is available, malnutrition is very likely to occur.

## CONCLUSIONS

### *Continuum of Casualty*

The information in the preceding pages provides compelling evidence that malnutrition can prevent the expression of the full genetic potential for mental development. The definitive study to prove causality and to study the significance of timing, environmental compensation, and so on, has not been done; it cannot be done because of the ethical restrictions on experimentation with human subjects.

Nutrition begins to affect the life of a baby long before it is born. The intergenerational data from the Aberdeen studies (Baird et al., 1953; Baird et al., 1954) demonstrated that the nutrition of the baby's mother, grandmother, and perhaps other ancestors can influence his own chances for unimpeded physical and mental development. One mechanism by which the intergenerational factor expresses its influence appears to be via the mother's stature, particularly the structure of her pelvis. Smaller women have an increased risk of giving birth to injured or low birth weight babies.

Another way in which nutrition affects a child prenatally is through intrauterine environment. In multiple births it may be the case that one fetus receives less arterial nourishment than the other. The deprived fetus may be stunted in physical or mental growth. In single births a fetus can still suffer from intrauterine malnutrition. An obvious factor that can affect the mother's consumption of crucial nutrients is income. Clearly, it is more difficult to provide adequately nutritious diets on poverty-level incomes than on middle-class incomes.

The extent of malnutrition in this country is directly correlated with income level (Citizens' Board, 1968). This relationship is easy to understand. In Massachusetts, as one example, welfare recipients receive approximately 30¢ per person, per meal (Massachusetts Welfare Rights Organization, 1969). Neither this amount nor the surplus commodity food available can provide adequate amounts of essential nutrients, especially since high protein foods are always more expensive than starches and other carbohydrates.

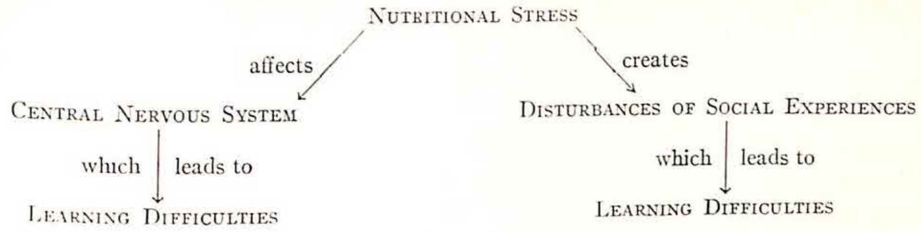


Fig. 1. Model of the effects of nutritional stress.

After a baby is born he is still vulnerable to the effects of nutritional deficiencies, some of which are reversible. Kugelmass, Poull, and Samuel (1944) found they were able to increase the average IQ score of a group of malnourished children by 18 points, by increasing their intake of nutrients. The IQs of mentally retarded children increased by an average of 10 points with the same treatment. The timing of the rehabilitation was crucial: the earlier the treatment, the greater was its effect.

With regard to the factor of timing, perhaps it is most helpful to imagine a continuum of casualty, to borrow the term used by Knobloch and Pasamanick. Beginning 6 months before birth and lasting until 6 months after birth, cell division is providing rapid growth of brain tissue. Nutritional deprivation during this period can severely retard psychological development, and it is probably irreversible unless treatment begins before cell division ends (Cravioto & Robles, 1965; Winick, 1968). From 6 months through the first or second year of life the rate of brain growth is still very rapid, mainly by protein synthesis, but nutritional rehabilitation still can be partially effective (Stoch & Smythe, 1967). Beyond this point, the rate of brain growth is significantly slower, and Kugelmass et al. (1944) found negligible effects of treatment on IQ scores beyond the age of 4 years.

### *The Cravioto-Birch Model: Interaction of Organic and Social Factors*

The preceding discussion has considered only the organic effects of malnutrition, and how organic changes might retard mental development. Social factors are also important. As usual, the question should not be phrased in terms of one or the other type of influence, but rather the interaction of both. Figure 1 is a model of the effects of nutritional stress, as proposed by Cravioto and Birch (Birch, 1968; see Footnote 3; Cravioto, 1966), which illustrates the ways in which social learning factors and organic changes probably interact. The left-hand chain illustrates the sequence of events this paper has emphasized. Nutritional deprivation (either pre- or postnatal) leads to changes in the brain and impedes mental development. Learning difficulties, or, in the extreme case, mental deficiency, result.

The other chain illustrates the fact that nutritional deprivation can work indirectly on mental capacity. As Cravioto[4] explained, malnutrition can interfere with learning by three indirect mechanisms:

*Loss of learning time.* Children who receive insufficient nutrients are more vulnerable to illness and consequently miss more school days than healthy children.

*Changes in concentration or motivation.* Especially for children who are not only malnourished, but also hungry, continual distraction by the discomforts of an empty stomach can seriously interfere with the child's efforts in school.

*Loss of the social interaction involved in meal times.* As Cravioto pointed out, we do not often think of the amount of learning that occurs at most family dinner tables: learning of symbolic and social systems, learning of cultural norms, and exposure perhaps to adult interests. With poverty and lack of food, this social learning setting may break down, and the malnourished child is deprived of still another type of educational experience.

This model illustrates the complexity of the problem of malnutrition and mental development. Birch (1968) summarized:

---

[4] J. Cravioto. Symposium presented at the meeting of the American Association for the Advancement of Science, Boston, December 1969.

> Children who are ill-nourished are reduced in their responsiveness to the environment, distracted by their visceral state, and reduced in their ability to progress and endure in learning conditions. . . . Consequently . . . [there is] a reduction in the profit which a child may derive from exposure to opportunities for experience [p. 596].

### *Urgency of the Problem*

Preschool children appear to be the major at-risk group all over the world. According to Béhar (1968), approximately 75% of preschool children in South America, Asia, and Africa are underweight for their age. Since about 400 million children under 5 years of age live in these areas, he suggested that there are approximately 300 million preschool children—60% of the total preschool population of the world—suffering from mild to moderate protein-calorie malnutrition.

Malnutrition is also an urgent problem in the United States. In April 1968 the Citizens' Board of Inquiry into Hunger and Malnutrition in the United States published its report, which claimed that more than 10 million Americans were affected by hunger and malnutrition. Congress, in the "Partnership for Health" Bill, charged the Secretary of Health, Education and Welfare (HEW) to conduct a survey of the "prevalence and location of serious hunger and malnutrition, and the health problems incident thereto, in the U. S. [Schaffer, 1968, p. 1130]." The survey was conducted by the Nutrition Program of the Public Health Service. In the spring of 1968 HEW published its report, which stated:

> children are smaller, they suffer from anemia and the effects of substandard protein and vitamin dietary intake. There is reason to believe that the continuing levels of inadequacy in dietary intake are associated with physical and mental damage . . . [Nelson, 1968, p. 1434].

In 1969 the White House Conference on Food, Nutrition and Health presented the President with a 625-page report. The priority list for target groups reflects some of the issues discussed in this paper: preschool children, pregnant women, primary school children, and then other specific deprived groups, such as American Indians and migrant workers. The recommendations of the conference, if implemented immediately, would apparently eliminate hunger and malnutrition within 3 years. When Senate subcommittee hearings were held in February 1971, it became obvious that in reality little progress had been made since the White House Conference.

The issue of malnutrition has traditionally been considered the province of medical and public health experts, as demonstrated by the bibliographic references for this review. It is puzzling that psychologists have shown so little interest in the psychological effects of malnutrition. (A notable exception is the Bay Area Psychologists and Social Scientists for Social Action; see Opton, Margen, Neilands, Pilisuk, & Sanford.[5]) There is much yet to be learned about motor deficits, personality development, and learning defects and their possible relationships to malnutrition.[6]

---

[5] E. M. Opton, S. Margen, J. B. Neilands, M. Pilisuk, & N. Sanford. Hunger causes mental retardation: A Report to the California Assembly Committee on Health and Welfare, 1969. (Mimeo)

[6] Research published since this review was submitted (Kotch, 1970; Kotz, 1969; Levitsky & Barnes, 1972; Rosso & Winick, 1970; Van Marthens & Zamenhof, 1969; Winick, 1970; Zamenhof, Van Marthens, & Grauel, 1971) includes several excellent review articles, new evidence on the intergenerational effect and on environmental interaction, and also a thorough analysis of the political obstacles to eliminating hunger in the United States."

## REFERENCES

ANDERSON, W. J., BAIRD, D., & THOMSON, A. M. Epidemiology of stillbirths and infant deaths due to congenital malformation. *Lancet,* 1958, **1**, 1304–1306.

ANTONOV, A. N. Children born during the siege of Leningrad in 1942. *Journal of Pediatrics,* 1947, **30**, 250–259.

BABSON, S., KANGAS, J., YOUNG, N., & BRAMHALL, J. Growth and development of twins of dissimilar size at birth. *Pediatrics,* 1964, **33**, 327–333.

BAIRD, D., THOMSON, A. M., & DUNCAN, E. H. L. Causes and prevention of stillbirths and first week deaths, II: Evidence from Aberdeen clinical records. *Journal of Obstetrics and Gynaecology of the British Empire,* 1953, **60**, 17–30.

BAIRD, D., WALKER, J., & THOMSON, A. M. Causes and prevention of stillbirths and first week deaths, III: Classification of deaths by clinical cause: Effect of age, parity, and length of gestation on death rates by causes. *Journal of Obstetrics and Gynaecology of the British Empire,* 1954, **61**, 433–448.

BÉHAR, M. Prevalence of malnutrition among preschool children of developing countries. In N. W. Scrimshaw & J. E. Gordon (Eds.), *Malnutrition, learning and behavior*. Cambridge, Mass.: M.I.T. Press, 1968.

BIRCH, H. G. Health and education of socially disadvantaged children. *Developmental Medicine and Child Neurology,* 1968, **10**, 580–599.

BIRCH, H. G., & CRAVIOTO, J. Infection, nutrition and environment in mental development. In H. F. Eichenwald (Ed.), *The prevention of mental retardation through control of infectious diseases.* (Public Health Service Pub. No. 1692) Washington, D. C.: United States Government Printing Office, 1968.

BROWN, R. E. Organ weight in malnutrition with special reference to brain weight. *Developmental Medicine and Child Neurology,* 1966, **8**, 512–522.

BURKE, B. S., HARDING, V. V., & STUART, H. C. Nutrition studies during pregnancy: 4. Relation of protein content of mother's diet during pregnancy to birth length, birth weight, and condition of infant at birth. *Journal of Pediatrics,* 1943, **23**, 506.

CABAK, V., & NAJDANVIC, R. Effect of undernutrition in early life on physical and mental development. *Archives of Disease in Childhood,* 1965, **40**, 532–534.

CALDWELL, D. F., & CHURCHILL, J. A. Learning ability in the progeny of rats administered a protein-deficient diet during the second half of gestation. *Neurology,* 1967, **17**, 95–99.

CHASE, H. P., & MARTIN, H. P. Undernutrition and child development. *New England Journal of Medicine,* 1970, **282**, 933–939.

CHATTERJEE, K. K., & MUKHERJEE, K. L. Phospholipids of the liver in children suffering from protein-calorie undernutrition. *British Journal of Nutrition,* 1968, **22**, 145–151.

CHRISTAKIS, G., MIRIDJANIAN, A., NATH, L., KHURANA, H., COWELL, C., ARCHER, M., FRANK, O., ZIFFER, H., BAKER, H., & JAMES, G. A nutritional epidemiologic investigation of 642 New York City children. *American Journal of Clinical Nutrition,* 1968, **21**, 107–126.

CHURCHILL, J. A. The relationship between intelligence and birth weight in twins. *Neurology,* 1965, **15**, 341–347.

CHURCHILL, J. A., NEFF, J. W., & CALDWELL, D. F. Birth weight and intelligence. *Obstetrics and Gynecology,* 1966, **28**, 425–429.

CITIZEN'S BOARD OF INQUIRY INTO HUNGER AND MALNUTRITION IN THE UNITED STATES. *Hunger, U.S.A.* Boston: Beacon Press, 1968.

CRAVIOTO, J. Malnutrition and behavioral development in the pre-school child. In, *Pre-school child malnutrition.* (Publ. No. 1282) Washington, D. C.: National Academy of Science-National Research Council, 1966.

CRAVIOTO, J., & DELICARDIE, E. R. Intersensory development of school-age children. In N. W. Scrimshaw & J. E. Gordon (Eds.), *Malnutrition, learning and behavior*. Cambridge, Mass.: M.I.T. Press, 1968.

CRAVIOTO, J., LICARDIE, E. R., MONTIEL, R., & BIRCH, H. G. Motor and adaptive development of premature infants from a preindustrial setting during the first year of life. *Biologia Neonatorum,* 1967, **11**, 151–158.

CRAVIOTO, J., & ROBLES, B. Evolution of adaptive and motor behavior during rehabilitation from kwashiorkor. *American Journal of Orthopsychiatry,* 1965, **35**, 449–464.

CROME, L. C., & STERN, J. *The pathology of mental retardation.* London: Churchill, 1967.

DAVISON, A. N., & DOBBING, J. Myelination as a vulnerable period in brain development. *British Medical Bulletin,* 1966, **22**, 40–44.

DICKERSON, J. W., DOBBING, J., & MCCANCE, R. A. The effect of undernutrition on the postnatal development of the brain and cord in pigs. *Proceedings of the Royal Society of London,* 1967, **166**, 396.

DIECKMANN, W. J., TURNER, D. F., MEILLER, E. J., SAVAGE, L. J., HILL, A. J., STRAUBE, M. T., POTTINGER, R. E., & RYNKIEWICZ, L. M. Observations on protein intake and the health of the mother and baby: I. Clinical and laboratory findings. *Journal of the American Dietetic Association,* 1951, **27**, 1046.

DRISCOLL, S. G. Why are twins dissimilar? *Pediatrics,* 1964, **33**, 325–326.

EBBS, J. H., BROWN, A., TISDALL, F. F., MOYLE, W. J., & BELL, M. The influence of improved prenatal nutrition upon the infant. *Canadian Medical Association Journal,* 1942, **46**, 6–8.

GOPALAN, C. Malnutrition in childhood in the tropics. *British Medical Journal,* 1967, **4**, 603–607.

HARRELL, R. F., WOODYARD, E., & GATES, A. I. *The effects of mothers' diets on the intelligence of offspring.* New York: Teachers College, 1955.

JEANS, P. C., SMITH, M. B., & STEARNS, G. Incidence of prematurity in relation to maternal nutrition. *Journal of the American Dietetic Association,* 1955, **31**, 576–581.

JELLIFFE, D. B. *Child nutrition in developing countries.* Washington, D. C.: United States Government Printing Office, 1968.

KIRKWOOD, W. Aspects of fetal environment. In H. Wolff (Ed.), *Mechanisms of congenital malformation.* New York: Association for the Aid of Crippled Children, 1955.

KLIPSTEIN, F. A. Nutritional anemia among America's poor. *Annals of Internal Medicine,* 1968, **68**, 1125–1127.

KNOBLOCH, H., RIDER, R., HARPER, P., & PASAMANICK, B. Neuropsychiatric sequelae of prematurity: A longitudinal study. *Journal of the American Medical Association,* 1956, **161**, 581–585.

KOTCH, J. Protein-calorie malnutrition and mental retardation. *Social Science and Medicine,* 1970, **4**, 629–644.

KOTZ, N. Let them eat promises: The politics of hunger in America. New York: Doubleday & Co., 1969.

Kugelmass, J. N., Poull, L. E., & Samuel, E. L. Nutritional improvement of child mentality. *American Journal of the Medical Sciences,* 1944, **208**, 631–633.

Laporte, M. The effect of war imposed dietary limitations on growth of Paris school children. *American Journal of Disease in Children,* 1946, **71**, 244.

Levitsky, D. A., & Barnes, R. H. Nutritional and environmental interactions in the behavioral development of the rat: Long-term effects. *Science,* 1972, **176**, 68–71.

Masland, R. L. The prevention of mental subnormality. In R. L. Masland, S. B. Sarason, & T. Gladwin, *Mental subnormality*. New York: Basic Books, 1958.

Massachusetts Welfare Rights Organization. Five lies about welfare. Cambridge, Mass.: Author, 1969.

McCance, R. A., Widdowson, E. M., & Verdon-Roe, C. M. A study of English diets by the individual method: III. Pregnant women at different economic levels. *Journal of Hygiene,* 1938, **38**, 596.

McGanity, W. J., Bridgforth, E. B., & Darby, W. J. Effect of reproductive cycle on nutritional status and requirements. *Journal of the American Medical Association,* 1958, **168**, 2138–2145.

Montagu, M. F. A. *Prenatal influences*. Springfield, Ill.: Charles C Thomas, 1962.

Naeye, R. L., Diener, M. M., & Dellinger, W. S. Urban poverty: Effects on prenatal nutrition. *Science,* 1969, **166**, 1206.

Nelson, B. Hunger and malnutrition: HEW says nation must know more and do more. *Science,* 1968, **160**, 1433–1436.

Palmer, C. E. Height and weight of children of the depression poor. *Public Health Reports,* 1935, **59**, 33.

Parekh, U. C., Pherwani, A., Udani, P. M., & Mukherjee, S. Brain weight and head circumference in fetus, infant and children of different nutritional and socio-economic groups. *Indian Pediatrics,* 1970, **7**, 347–358.

Pasamanick, B., & Lilienfeld, A. Association of maternal and fetal factors with the development of mental deficiency: I. Abnormalities in the prenatal and paranatal periods. *Journal of the American Medical Association,* 1955, **159**, 155–160.

Platt, B. S., Heard, R. C., & Stewart, R. J. Experimental protein-calorie deficiency. In H. N. Munro & J. B. Allison (Eds.), *Mammalian protein metabolism*. New York: Academic Press, 1964.

Pollitt, E., & Granoff, D. Mental and motor development of Peruvian children treated for severe malnutrition. *Revista Interamericana de Psicologia,* 1967, **1**, 93–102. (*Psychological Abstracts,* 1967, **41**, No. 1169)

Robinson, H. B., & Robinson, N. M. *The mentally retarded child*. New York: McGraw-Hill, 1965.

Rosso, P., Hormazábel, J., & Winick, M. Changes in brain weight, cholesterol, phospholipid, and DNA content in marasmic children. *American Journal of Clinical Nutrition,* 1970, **23**, 1275–1279.

Schaffer, A. E. The national nutrition survey. *Annals of Internal Medicine,* 1968, **68**, 1130–1131.

Scrimshaw, N. W., & Gordon, J. E. (Eds.) *Malnutrition, learning and behavior*. Cambridge, Mass.: M.I.T. Press, 1968.

Stoch, M. B., & Smythe, P. M. The effect of undernutrition during infancy on subsequent brain growth and intellectual development. *South African Medical Journal,* 1967, **41**, 1027.

Strauss, M. B. Anemia of infancy from maternal iron deficiency in pregnancy. *Journal of Clinical Investigation,* 1953, **12**, 345–353.

Thomson, A. M. Maternal stature and reproductive efficiency. *Eugenics Review,* 1959, **51**, 157–162.

Thomson, A. M. Historical perspectives of nutrition, reproduction, and growth. In N. W. Scrimshaw & J. E. Gordon (Eds.), *Malnutrition, learning and behavior*. Cambridge, Mass.: M. I. T. Press, 1968.

Toverud, K. U., Stearns, G., & Macy, I. G. *Maternal nutrition and child health: An interpretive review*. (Bulletin No. 123) Washington, D. C.: National Research Council, 1950.

Van Marthens, E., & Zamenhof, S. Deoxyribonucleic acid of neonatal rat cerebrum increased by operative restriction of litter size. *Experimental Neurology,* 1969, **23**, 214–219.

Walker, J. Aspects of fetal environment. In H. Wolff (Ed.), *Mechanisms of congenital malformation*. New York: Association for the Aid of Crippled Children, 1955.

Willerman, L., & Churchill, J. A. Intelligence and birth weight in identical twins. *Child Development,* 1967, **38**, 623–629.

Winick, M. Nutrition and cell growth. *Nutrition Reviews,* 1968, **26**, 195–197.

Winick, M. Nutrition and mental development. *Medical Clinics of North America,* 1970, **54**, 1413–1428. (a)

Winick, M. Nutrition and nerve cell growth. *Federation Proceedings,* 1970, **29**, 1510–1515. (b)

Winick, M., & Noble, A. Cellular response in rats during malnutrition at various ages. *Journal of Nutrition,* 1966, **89**, 300–306.

Winick, M., Rosso, P., & Waterlow, J. Cellular growth of cerebrum, cerebellum, and brain stem in normal and marasmic children. *Experimental Neurology,* 1970, **26**, 393–400.

Zamenhof, S., Van Marthens, E., & Grauel, L. DNA (cell number) in neonatal brain: Second generation ($F_2$) alteration by maternal ($F_0$) dietary protein restriction. *Science,* 1971, **172**, 850–851.

(Received March 10, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 5, 335–350

# TASK STRUCTURE AND THE EXPERIMENTAL MODIFICATION OF VERBAL BEHAVIOR

G. ALAN MARLATT[1]

*University of Wisconsin*

Numerous studies have investigated techniques assumed to modify verbal behavior. Attention is focused on three of the most important procedures: instructions, verbal models, and verbal conditioning. These procedures are reviewed and interpreted as sources of information available to the subject in the experimental task. The effects of task structure and subject uncertainty are discussed in terms of these sources of information and their effects upon the subject's awareness and intention related to task performance. Studies that have compared the use of the three modification techniques with verbal behavior are evaluated within the context of this theoretical approach. Finally, the similarities and differences between experimental studies of verbal behavior and the psychotherapeutic setting are discussed in terms of applying modification procedures to problems outside of the laboratory.

Numerous studies have attempted to modify the verbal behavior of subjects in a variety of experimental tasks. Among the modification procedures that have been investigated in this regard, three techniques have received the greatest amount of attention: use of direct instructions to elicit verbal change, administration of verbal reinforcement in an attempt to "condition" emitted verbal responses, and the provision of verbal models to evoke change mediated by principles of observational learning.

All of the various modification procedures share a common feature. The agent of modification, the experimenter, selects a target verbal state (some change in the frequency or type of verbal response) which he attempts to produce in the individual subject. Producing the target state by the selective application of a certain "treatment" (modification technique) is the *purpose* of such an experiment, as construed by the experimenter. The subject, upon first entering the laboratory, usually does not know the true purpose of the experiment until so informed by the experimenter (often at the completion of the study, when the subject is told the "real" purpose). As such, the subject can be described as in an initial state of uncertainty about his expected performance in the experimental task.

As described in this context, the subject is uncertain as to how he should perform successfully in the task. The criteria of success can only be obtained through knowledge of the purpose of the experiment. The experimenter reduces the subject's uncertainty by providing him with several *sources of information*. In the scope of the present paper, these sources may be explicit in the form of instructions, verbal reinforcement, or presentation of models. Additional sources of information are implicit in the nature of the task itself (information provided by the degree of task structure or ambiguity), or are transmitted in subtle form to the subject by the experimenter's behavior. Attention must be paid to such factors as the purpose of the experiment (as construed by experimenter and subject), and how information provided to the subject affects his performance in the experimental task, if one is to discuss how experimental findings are to be applied in settings outside the laboratory.

The purpose of this paper is to provide a theoretical framework that attempts to integrate existing knowledge about procedures used to modify verbal behavior. Specifically, an attempt is made to describe and analyze instructions, verbal conditioning, and the use of models as sources of information available to the subject. The ways in which this information is utilized (in terms of the subject's awareness and intentions in the experiment) are discussed as possible determinants of overt task performance. In addition, certain similarities and differences between the laboratory setting and traditional forms of verbal psychotherapy are examined to assess the potential effectiveness of applying modification procedures to problems outside of the laboratory.

---

[1] The author is especially grateful to A. Bandura, F. H. Kanfer, L. H. Levy, and I. G. Sarason for reading earlier versions of this paper and providing helpful comments.

Requests for reprints should be sent to G. Alan Marlatt, who is now at the Department of Psychology, University of Washington, Seattle, Washington 98195.

## TASK AMBIGUITY AND SUBJECT UNCERTAINTY

For the purposes of this discussion, *ambiguity* can be said to be a property of a task when two or more equally likely response alternatives are available to the subject. "Equally likely" alternatives exist when no information is provided (by the task structure itself or by the experimenter) as to which alternative is more appropriate in terms of the purpose of the experiment, and when response consequences are equal or equally unknown to the subject. It can be said that the subject, when faced with an ambiguous task, is in a state of response *uncertainty*.

Any information provided to the subject may serve as discriminative stimuli that enable him to discriminate among the available alternatives and select one that is congruent with the experimental purpose. Selection of this response reduces the subject's uncertainty and is one factor that facilitates performance in the task. Sources of information that may be available to the subject include verbal directives or elicitations given by the experimenter or interviewer prior to or during the task itself (instructions), prior observation of the performance of another individual in the same task (modeling), or response-contingent "feedback" during the subject's performance (as in verbal conditioning studies).

It is with regard to the role of verbal reinforcement that controversy arises in the informational analysis described above. The argument centers around the question of whether such reinforcement primarily provides the subject with information that can be used to reduce response uncertainty (cognitive orientation), or provides an associative strengthening of the response by means of factors such as the incentive or affective quality of the reinforcement (behavioral orientation). This important issue can be further explored by examining the ways in which a subject may use the sources of information available to him. Although the argument usually focuses upon the disputed role of "awareness" as it affects changes in verbal behavior, this discussion also includes intention as an additional cognitive construct.

## AWARENESS AND INTENTION

The role of awareness in the process of learning is a topic of considerable controversy (Adams, 1957; Eriksen, 1960). In the realm of studies dealing with verbal behavior, the controversy appears central in interpreting performance in verbal conditioning tasks. As considerable attention has been paid to this topic by other authors (e.g., Das, 1969; Eriksen, 1962; Kanfer, 1968), only the highlights of the argument are presented here.

There are two main theoretical viewpoints pertaining to awareness in the process of verbal conditioning. The behavioral (or nonmediational) view holds that awareness is a resultant of changes in behavior and can be considered as a dependent variable. Response changes in the task and subsequent changes in awareness are both considered to be a function of manipulations at the independent variable level. This orientation, with minor changes, is held by Kanfer (1968), Krasner (1962), and Skinner (1957), among others. Cognitive theorists, on the other hand, believe that awareness is a necessary condition for response changes to occur, and posit awareness as an independent or intervening variable in the control of the verbal response (e.g., Dulany, 1962, 1968; Spielberger, 1965). Intermediate or mixed-model theories have also been put forth as explanations of awareness (e.g., Farber, 1963; Verplanck, 1962).

### *Awareness of What?*

The issue of awareness in verbal conditioning studies is further complicated by the confusion that exists as to the definition of this term. Eriksen (1962) has defined awareness in this context as the ability of the subject to verbalize correctly the relationship between an operant verbal response and the contingent reinforcer. The ability to verbalize in this way is a *learned* phenomenon in itself, and is similar to the learning of any rule or construct. Dulany (1962) has also attempted to categorize what he sees to be differing levels of awareness in the verbal conditioning task.

It is the position of the present author that the concept of awareness in verbal conditioning studies can be expanded to refer to the degree of knowledge possessed by the subject as to the true (experimenter's) purpose of the experiment. The experimenter's purpose is to produce changes in a delineated response class through the application of specified modification techniques. Strictly speaking, as the extent of awareness in the experimental setting is determined on the basis of criteria established by the experimenter, the concept of awareness should be defined as the degree of congruence between the subject's own perception of the purpose of the experiment and the experimenter's purpose. The subject's awareness, then, may vary along a continuum of "correctness" as determined by external standards. Although this view is consistent with the definitions provided by Eriksen and Dulany, it is also an extension of these positions in that awareness can be developed as a function of *any* sources of information in addition to reinforcement available to the subject in the task.

*Awareness and task structure.* Support for a broader definition of the concept of awareness derives from an examination of the results of studies that have found a relationship between ambiguity of the verbal conditioning task and the degree of awareness reported by subjects. A review of such studies reveals that increased awareness occurs as a function of the greater availability of information as tasks become more structured and less ambiguous to the subject. Information about the task has proved to facilitate awareness in a number of studies (Kanfer & Marston, 1961; Kanfer & McBrearty, 1961; Krasner & Ullmann, 1963; Tatz, 1960). Comparison of the proportion of aware subjects across a variety of conditioning tasks would be of additional value in this respect.

Because of the highly structured nature of verbal conditioning tasks such as the Taffel (1955) sentence-construction procedure, demand characteristics that determine the subject's awareness of the experimental purpose may become more effective than they would be in a more ambiguous task. This interpretation is clouded, however, because information provided by the task structure increases with the degree of task simplicity itself, and may lead to other nontask demand characteristics becoming redundant. The subject's chances of being correct in his hypotheses increase as the number of possible correct alternatives decreases, as in a relatively simple task. As Bandura (1969) has noted:

> When the critical response class is unambiguous and the response alternatives are severely curtailed, both awareness and "learning" are most likely to occur as a one-trial event rather than as an incremental process [p. 575].

A similar argument can be made for experimenter bias effects in relation to task complexity. Subjects exposed to a highly ambiguous task may respond in a greater fashion to extra-task cues provided by both demand characteristics and experimenter bias, because of the relatively high state of subject uncertainty. Conversely, in a highly structured task, such as the Taffel conditioning procedure, experimenter bias effects may prove to be ineffective due to the sufficient information provided by the interaction of task structure and reinforcement (cf. Zegers, 1968).

### *Intentions*

Awareness, as discussed in the present context, refers to a state of knowledge. This knowledge, pertaining to the subject's perception of the experimenter's purpose, is in itself a static cognitive state. It represents "what is learned" in the experimental situation. The only observable manifestations of this learning are the subject's overt task performance and his statements about the determinants of his performance.

Awareness itself, however, is not sufficient to account for overt task performance. There must be some incentive present (such as desire to please the experimenter, need for confirmation of one's hypotheses, wanting to receive social approval, etc.) that activates the subject's performance. This incentive has the effect of producing an *intention* to act. Intention is here defined as a set of self-generated instructions based on both the subject's awareness of the "correct" response, and sufficient incentive conditions to motivate potential performance.

The definition of intention presented above is similar to Dulany's (1962) discussion of "behavioral intention." The concept of intention is important from a cognitive point of view because it provides an index of awareness in the experimental task. It is here assumed that *learning* can occur in the absence of immediate performance of the task response, and that information about what has been learned can be ascertained by asking subjects to verbalize their intentions.

It is important to note that awareness and intention may not always be congruent. Knowledge of the experimental purpose (awareness) may lead to the intent to perform the "correct" response, or (depending on particular incentive conditions) may lead to a different intent, such as to inhibit responding or to produce a different or "opposite" response. Farber (1963) has discussed a verbal conditioning study in which subjects were assessed for awareness in posttask inquiry. He found that aware subjects either showed performance that could be considered congruent with their intent to conform to the purpose of the experiment, or showed performance that could be considered incongruent with conforming intentions (his "awarenonforming" subjects). Similar results have been reported by Holmes (1967). The relation between awareness and intention has been stated by Dulany (1962): "That a human subject does what he thinks he is supposed to do [awareness] if he wants to [intention] . . . [p. 109]."

The behavioral point of view as stated by Verplanck (1962) also assumes the independence of awareness and intentions, but adds that intentions and performance are additional independent response systems, independently subject to reinforcement control: "Experimentally, the subject's rules, his 'hypotheses,' can be dissociated to a degree from the behaviors that they are presumed to direct. He does not carry out his intentions [Verplanck, 1962, p. 138]."

Another possible behavioral stance would be that the subject does learn without awareness, and that he *subsequently* attributes his behavioral changes to his own intentions. In order to clarify the controversy that exists concerning awareness and intention, a discussion of possible methodological strategies that might be employed to assess these two factors would be in order.

### *Assessment of Awareness and Intentions*

From the behavioral standpoint, factors such as awareness and intentions are artifacts, nuisance variables that are to be controlled for, if possible. Kanfer (1968), for example, has suggested the use of extensive pretraining procedures that would produce habituation of individual differences and behavioral "fluctuations" of this type. Another controlling procedure has been used by Oakes (Dixon & Oakes, 1965; Oakes, 1967), who instructed subjects to engage in a distracting task (such as naming colors) between verbal conditioning trials. It is assumed that performance of this intertrial task will interfere with the formulation and rehearsal of awareness hypotheses.

*Criticisms of awareness assessment procedures.* When awareness is assessed in studies of verbal behavior, the usual procedure is to question the subject *following* his performance in the task. Subjects are then assigned to various categories of awareness on the basis of their statements about the experiment (e.g., Spielberger, 1962). Considerable criticism of this procedure has been put forth (Kanfer, 1968; Spielberger & DeNike, 1966). One main criticism has been that this method cannot determine the point in the task sequence where awareness first develops. Does awareness occur prior to, concurrently with, or subsequent to changes in overt performance? Another possibility is that reports of awareness are determined by the sensitizing nature of the inquiry procedure itself; several attempts have been made to construct assessment interviews which carefully proceed from general to specific questions (e.g., Spielberger & Levin, 1962). Attempts to overcome certain of the difficulties associated with posttask assessment of awareness have been made by questioning subjects at specified periods during the task performance itself (e.g., DeNike, 1964; Kennedy, 1970; Spielberger, Bernstein, & Ratliff, 1966).

A second body of criticism directed at the typical assessment procedure concerns the conditions under which subjects would be likely to report awareness to the experimenter. The argument is that self-reports of awareness are greatly affected by the demand characteristics of the experiment (as opposed to "pure" awareness of response contingencies in the absence of the demand incentives to discover them). This point of view (cf. Marx, 1958) loses its impact in terms of the definition of awareness as presented in this paper, as it is assumed that *all* extra sources of information (including demand characteristics), in addition to response-reinforcement contingencies, combine to determine awareness of the experimental purpose. Demand characteristics per se will thus have equal effects on the development of awareness and the subsequent "eagerness" of subjects to report their hypotheses.

*Manipulating awareness*. One alternative method to determine the effects of awareness is to manipulate awareness levels *prior* to task performance. If subjects were provided information as to the "true" experimental purpose in the form of instructions, or by being "tipped off" by a confederate subject (Levy, 1967a), it could be postulated that the role of reinforcement as a conveyor of *information* as to the purpose of the study would be considerably less important. For example, if subjects were told that the purpose of the study was to say nouns in three selected topic areas (topics determined by independent assessment to have a relatively high operant base rate), with no particular restrictions as to frequency or order within topic categories, would reinforcement applied to one particular category produce a "conditioning" effect compared to the performance of subjects not provided with prior information as to the purpose?

*Assessment of intentions*. Another possible solution to the general problem of the relationship between awareness and behavioral change is to assess the subject's intentions. Verbal reports of intentions to perform can be ascertained in the *absence* of actual task behavior. Subjects who are given information about the experiment, in the form of instructions or by observing the behavior of another subject who participates in the task, could be asked to state their intended behavior in the same task. The assessment of intentions could occur at various stages of the experimental sequence, in order to determine the temporal development of awareness hypotheses. Subjects could be queried about their expectations, awareness of experimental purpose, and intentions at fixed intervals prior to, during, and following the experiment.

## SOURCES OF INFORMATION

Instructions, the use of models, and the application of reinforcement can be considered as the three most important techniques used to experimentally manipulate verbal behavior. As conceived within the present theoretical context, these techniques share a common property in that they all serve as sources of information to the subject in the experimental setting.

### *Instructions*

Considering that some form of instructions is used in practically every psychological experiment involving human subjects, it is surprising that so little attention has been paid to the definition and theoretical analysis of this factor. Part of the reason for this neglect has been that instructions have been used for such a wide variety of purposes in the experimental situation that a single, all-encompassing definition would be of little meaning. Gagne (1964), in a discussion of problem-solving behavior, has outlined some of the key functions of instructions, including (*a*) identification of the terminal (or criterion) performance in the task; (*b*) identification of salient parts of the stimulus situation; (*c*) aiding recall of relevant subordinate performance capabilities necessary to the task; and (*d*) helping to "channel" thinking, in terms of strengthening or weakening prior hypotheses of subjects. In addition to these attention-directing and performance specification functions, instructions have also been used to manipulate motivational levels (by such means as describing incentives that may be obtained during or following performance or by producing changes in "drive level," as in self-esteem, success–failure, and other similar manipulations). A recent paper by Sarason (in press) reviews evidence showing that personality factors such as anxiety may affect the extent to which individuals attend to and use motivational cues presented in the form of instructions.

In many of the studies involving verbal behavior as the dependent variable, a cognitively oriented definition would seem to apply. Instructions consist of explicit information (structure), provided by the experimenter to the subject, which has both instigational and directive functions. The instigational function of instructions is designed to initiate or facilitate intention to perform in a general sense (the "this is what we want you to do" aspect). The instigational nature of instructions is absent in other modification techniques, such as exposing observers to a model without prior instruction to copy the behavior. The directive function directs the subject's attention to stimulus conditions in the task and to specified performance responses required to engage in task behavior. It may also be designed to control extraneous behaviors by directing subjects *not* to engage in certain responses. Whereas the instigational component chiefly affects the subject's intentions to perform, as by manipulation of incentive conditions, the directive component contains information that may contribute to the subject's awareness of the experimental purpose. As a whole, instructions provide the subjects with the experimenter's *rules* that apply to the experiment. The subject can make use of these rules as mediators that guide his performance in the task.

A parallel definition could be stated from a more behavioral point of view. The informational quality of instructions can be described in terms of discriminative stimulus properties (Simkins, 1963). As noted by DiVesta and Walls (1967), it is necessary to assume that a subject's response to instructions depends on his prior reinforcement history for responding to instructions of a similar type. A history of positive reinforcement for following instructions would thus facilitate future adherence.

Instructions can be used to vary the ambiguity of the task by increasing or decreasing the number of response alternatives open to the subject. They can also be used to manipulate awareness, as defined in this paper. Awareness can be increased (and the subject's uncertainty decreased) by informing him of the actual or ostensible purpose of the experiment. Deceptive instructions usually are designed to control the subject's attempts to ascertain the "true" purpose when it is felt that such knowledge (or attempts to gain this knowledge) would confound the effects of other independent variables. Instructions in verbal conditioning experiments are frequently of this type, so as to "allow" the effects of verbal reinforcement to operate in a relatively pure form.

It is interesting to note that studies which have investigated human learning in the *absence* of instructions, in the area of "incidental learning" (cf. Postman, 1964), have been used to determine the role of intentions in the learning process. This approach also applies to the investigation of observational learning (modeling, identification), where direct instructions are usually absent. The understanding of the nature of instructions may be further advanced by a discussion of modeling procedures as they apply to the modification of verbal behavior.

### *Modeling*

It was stated that the main characteristic of instructions is that they provide the subject with a rule or principle, which can then be used as a response class mediator. During performance, instances of that response class are generated that are consistent with the rule given in the instructions. As such, performance mediated by instructions can be considered as a form of *deductive* reasoning. In contrast, provision of a model engaged in task performance is characterized by presentation of a series of examples, or response items within some preselected response class. In this case, the subject must derive a general rule, based on his observation of the model's performance, which would define the criteria of the response class(es) accounting for the instances demonstrated by the model. Derivation of a rule from observation of discrete examples can be considered as a type of *inductive* reasoning. Once the rule has been established, the observing subject must then deductively generate instances of the mediated response class during his own task performance.

By this definition, performance that is determined through the observation of a model's behavior is similar to the processes involved in concept formation (Hunt, 1962). Interpreting modeling in these terms is particularly useful in explaining changes in verbal behavior determined by observational learning procedures. To define verbal modeling effects as simply a matter of response matching or imitation fails to account for the diversity and generalization of performance that is found in behavior of this type. Subjects exposed to a verbal model do not just mimic identical response items in a one-to-one correspondence with the model's performance. Rather, they appear to also produce unique response items that fall within the general response class which the subject assumes to govern the model's performance.

*Modeling compared with instructions.* Recently there has been some discussion of the similarities between instructions and the use of models as determinants of behavior change. Bandura (1969), for example, has defined instructions as a form of "verbal modeling cues [p. 145]." Although he noted that instructions have both an instigational and modeling function, he appears to equate the two concerning their effects upon behavior.

> In studies comparing the relative efficacy of instructions and verbal modeling . . . , both types of influences produce their effects through verbal modeling and they differ only in the explicitness with which the required responses are defined [p. 146].[2]

In terms of the present discussion, instructions can be equated with modeling procedures only when the instructions are in exemplary form: that is, if they consist only of a series of examples within the target response class. In order to compare the effects of modeling and instructional techniques, it is necessary to independently vary the instigational and directive (provision of examples or a rule) components. It is, of course, possible to modify the modeling sequences by having the model state his own rule governing his performance in the task, and/or by instructing subjects to reproduce the model's behavior in their own task performance.

By defining modeling as the provision of examples within a response class in the absence of explicit instigations or exhortations to perform (as by instructing subjects to reproduce the behavior of the model), observational learning of this type falls within the category of incidental learning—learning in the absence of instructions to learn. The problem then becomes one of explaining the instigational aspects of observational learning. Two aspects seem of primary importance: factors determining whether or not the observer *attends* to the model's behavior, and factors determining the observer's subsequent *performance* of these behaviors. Various determinants of the model's salience, in terms of individual model and observer characteristics and situational factors, have been reviewed elsewhere (e.g., Flanders, 1968).

An additional determinant of the initiation of observational learning would seem to be the ambiguity of the performance task. Observers, faced with impending performance in a highly ambiguous task, are in a state of response uncertainty, and will seek out "signposts" as a source of direction. Provision of a model in these circumstances may increase the probability that the subject will attend to these cues. The effect of task ambiguity upon modeling behavior in an interview setting was examined in a recent study by Marlatt (1971).

While modeling in this context has been described as an example of incidental learning (in terms of the absence of explicit instructions), this does not imply that observers have no awareness or intention to perform in a manner similar to that of the model. Awareness and intention in the modeling situation would seem to bear the same relationship to task performance as they do in verbal conditioning, as demonstrated in a recent study by Hamilton, Thompson, and White (1970).

*Vicarious reinforcement.* Reinforcement to the model, as observed by the subject, has been defined as vicarious reinforcement (e.g., Kanfer, 1965). The function of vicarious reinforcement in the modification of verbal behavior would seem to be twofold. Reinforcement to the model may fulfill a directive function by directing the observer's attention to precise response items or sequences of the model's performance. In this sense, observation of reinforcement to the model may increase the observer's awareness of the experimenter's purpose. The informational quality of vicarious reinforcement would likely increase as a function of greater task ambiguity or the subject's uncertainty as to which response is appropriate in the experimental task. The directive role of evaluative "levels" of vicarious reinforcement (positive, negative, and neutral) has been shown to be an effective determinant of observers' performance in an ambiguous interview task (Marlatt, Jacobson, Johnson, & Morrice, 1970).

---

[2] Bandura (personal communication, April 1971) has also noted that "if [instructions] inform a person how to do something that he does not know, then the modeling function is prominent; if, on the other hand, they simply instruct a person to perform what he has already learned, then they serve mainly an instigational function."

A second function of vicarious reinforcement may involve incentives to perform the modeled behavior. Bandura (1969) has described this "affective role" of vicarious reinforcement in terms of the observer's anticipatory arousal to the sight of positive reinforcement and/or to the witnessing of the model's affective responses to reinforcement administration (possibly eliciting corresponding emotional cues in the observer, which become conditioned to the model, his behavior, or the task situation). The distinction between the informational and affective functions of reinforcement is discussed in the following section.

### *Verbal Conditioning*

In the study of verbal conditioning, controversy abounds, especially concerning the role of awareness as it affects (or is affected by) changes in verbal behavior. The core of the controversy lies with the fact that a verbal reinforcer, in the typical verbal conditioning task, has both informational and incentive (affective, motivational) properties. The dual nature of reinforcing stimulus has long been noted: Thorndike (1935), for example, discussed both the strengthening effect and informational influence of the "aftereffects of responses." The position that is argued in this paper is that the informational properties of a reinforcing event contribute to the subject's *awareness* (as defined earlier), and that the incentive properties primarily affect the subject's *intentions* regarding his performance in the task. A problem that continuously confronts an interpretation of this kind is that the single reinforcing stimulus is assumed to have both properties, and that it is impossible to separate out their respective effects (cf. Postman & Sassenrath, 1961). An attempt is made later in this section to suggest possible strategies that may provide further information on this issue.

*Reward training versus discriminated operant conditioning.* In attempting to delineate the informational effects of verbal reinforcement, it is of interest to compare the verbal operant response with an analogue in the animal laboratory, the operant bar press of the rat. The similarities between the two situations have been described by Kanfer (1968).

There are, however, a number of important differences between the two learning situations as well. The obvious difference is the human's capacity for speech. The capacity for speech also entails the capacity for verbally mediated awareness, for self-instructions in the form of intentions, and for the verbal interaction between experimenter and subject (instructions, verbal relationship statements, etc.). Equally important is that the verbal reinforcement is delivered in the context of an interpersonal relationship, as contrasted with the impersonal delivery of a food pellet to the rat in a Skinner box. Factors affecting the nature of the relationship between the subject and experimenter (social influence variables) may also determine aspects of the subject's performance.

The rat's bar-pressing performance has been described as an example of "reward training" (Grant, 1964), in which a reinforcement follows an emitted response. It may be that the verbal conditioning task can be better described as an instance of what Grant called "discriminated operant conditioning." In this form of training, the reinforcement follows a response *elicited* by a particular discriminative stimulus. It is possible to conceive of the verbal reinforcement "good" as having discriminative stimulus properties in this sense. The informational properties of verbal reinforcement have been stated by Skinner in a personal communication cited in a paper by Matarazzo, Saslow, and Pareis (1960):

> It seems to me quite reasonable that the so-called reinforcement . . . acts more as a discriminative stimulus than as a reinforcement. In talking to the lady on my right at a dinner party, I may try various subjects without much success. When I hit on something which gets a nod of interest, I continue along the same line, not so much because the response has been reinforced, as because the response plus the reinforcement act as an $S^D$ indicating that this is the kind of audience which reinforces certain kinds of responses.

Similarly, a nod or "mmm-hmm" from the experimenter to the subject in the verbal conditioning task may indicate to the subject that he is on the right track in terms of discovering the experimental purpose (awareness) and perhaps fulfilling it (intention).

A similar interpretation of the verbal conditioning task has been given by Insko and Cialdini (1969). In the "two-factor" theory presented by these authors, the verbal reinforcer is assumed to have two effects: an informational and a "motivating" effect. The latter is assumed to operate through the "creating of positive rapport or liking which serves to motivate conformity consistent with the conveyed information as to the interviewer's attitude [p. 334]." In other words, the reinforcer "good" tells the subject that the experimenter approves or likes the subject's response and, by implication, approves or likes the subject himself. Due to the process of reciprocation, the subject acquires some positive attraction for the experimenter that serves to motivate the conformity behavior. Evidence is cited by Cialdini and Insko (1969) in support of their theoretical position.

*Information and awareness.* If awareness is defined in terms of the subject's knowledge of the experimental purpose, administration of verbal reinforcement serves as just one source of information to the subject participating in a conditioning study. It would follow that if the subject were made aware of the purpose of the study through other sources than reinforcement administration, the information provided by reinforcement alone would become redundant. Subjects who have already become aware of the experimenter's purpose may not *attend* to reinforcement to the same degree as subjects who are uncertain as to the purpose. The ineffectiveness of reinforcement in a highly structured task has been demonstrated in a study by Weiss, Ekman, Ullman, and Krasner (1965).

Providing subjects with partial information as to the purpose of the verbal "conditioning" task may lead to effects that resemble typical acquisition curves even in the absence of reinforcement. Among the studies that have manipulated information available to subjects by the use of instructions, a paper by Di-Vesta and Walls (1967) demonstrated this type of finding. In a task involving choosing a response from verb pairs, subjects who were instructed that a "general principle" was involved and that they were to try to give as many words as possible conforming to this principle showed superior performance even in the absence of verbal reinforcement. The authors also found definite response tendencies (as opposed to random selection) during the operant base-line period; subsequent reinforcement was effective only if the reinforcement was congruent with the prior response tendency. It may be that prior instructions affect the informational value of reinforcement and that reinforcement in these instances may act as a source of confirmation when prior hypotheses (awareness) exist.

*Assessment of informational and incentive properties.* In order to understand further the relative contributions of the informational and incentive properties of verbal reinforcement, investigators should attempt to vary these qualities independently. Tasks varying in ambiguity could be compared, using the same reinforcement stimulus, in order to assess the relative effects of structure upon awareness ratings and performance (cf. Cairns, 1967). Tasks ranging from those high in structure (e.g., the Taffel task) to those low in structure (e.g., the unstructured self-disclosure interview) could be compared in this way. Similarly, by holding task structure constant, it would be possible to vary the incentive quality of the reinforcer in order to determine the effects upon intentions to perform in a manner congruent to the subject's awareness. It is possible to ask subjects to evaluate the reinforcing stimulus in terms of its "pull" value (Mandler & Kaplan, 1956); Spielberger, Levin, and Shepard (1962) found, for example, differences in performing as a function of how highly subjects rated the reinforcer ("good") in answer to the question, "Would you say you wanted me to say good?" In another study (Spielberger, Bernstein, & Ratliff, 1966) challenging subjects to obtain as many reinforcers ("mmm-hmm") as they could, produced facilitating incentive effects in aware subjects. A related strategy would be to vary the type of the reinforcement stimulus, ranging along a social–nonsocial dimension (Dulany, 1962; Finley & Staats, 1967). Verbal social reinforcements provided by the experimenter could be compared with stimuli such as tones, lights, and advances on a counter.

An additional strategy that could be used to investigate the incentive role of reinforcement would be to manipulate subject differences, holding task structure and type of reinforcement constant. Subjects could be assigned to groups in terms of such assessed differences as social desirability rating or dependency level (Alpert & Nakamura, 1968). Similar manipulations could be investigated in terms of the incentive components of instructions and modeling procedures. Reinforcement for certain task responses can be implied in the form of instructions given prior to performance; reinforcement to the model (vicarious reinforcement) could be studied in a similar fashion. Finally, experimenter characteristics could be varied in order to assess the effects of the relationship (positive or negative) between experimenter and subject upon response to reinforcement (Kessel & Barber, 1968). It has been found, for example (Ganzer & Sarason, 1964), that a hostile experimenter can produce an interaction effect with reinforcement administration.

### *Response Type*

In order to determine which procedure (instructions, modeling, or verbal reinforcement) would be appropriate for modifying verbal behavior in a particular situation, consideration should be paid to the type of target response selected. The informational properties of the modification technique will have different effects depending upon the ambiguity of the target response class, for example.

Controversy has arisen with responses that may be under some form of suppression (inhibition), such as emotional self-references or affect words. Bandura (1965) suggested that subjects instructed to give a particular verbal response should have no difficulty in doing so. Krasner (1965) has expressed disagreement with this point of view, saying:

> This is not as easy to do with a more ambiguous response class such as affect words. Here the response repertoire is under much less subject control. Even if his instructions were quite clear as to the task requirements he could not necessarily produce the desired operants [p. 218].

Instructions may be ineffective for this response class for other reasons as well: It is difficult to specify a set of rules that explicitly define vague response classes such as affect words; second, it may be very difficult to persuade subjects that the response consequences for admission of (say) negative self-references would not be aversive.

Problems also arise with the verbal conditioning approach. The subject may never make the response for which he is to be reinforced. In addition, because of the interpersonal nature of the verbal conditioning task (with verbal approval from the experimenter as the reinforcement), admission of negative affect statements may be seen as counterproductive for the subject bent on pleasing the experimenter. Difficulties in the conditioning of negative self-references have been noted in studies by Ganzer and Sarason (1964) and Merbaum and Lukens (1968). On the other hand, success has been reported for similar response classes by other investigators (cf. Hekmat & Lee, 1970; Kahn, 1967; Merbaum, 1963). Clarification of this issue and other problems associated with response type would call for an examination of studies that have compared various modification techniques.

### *Comparison of Modification Procedures*

*Instructions versus verbal conditioning.* Several studies have compared the relative effectiveness of instructions and verbal conditioning procedures. Following from the controversy described above, Merbaum and Lukens (1968) compared the effects of instructions, elicitations (direct questioning), and verbal reinforcement upon the frequency of positive and negative emotional words given by subjects in a task similar to the Thematic Apperception Test (telling stories about stick figures). For groups directed to give positive words, instructions were found to have the most facilitative effect, significantly greater than reinforcement alone (the reinforcement group was reliably superior to the control group, however). In the case of negative emotional words, instructions and elicitations were again more potent than reinforcement; no conditioning effect was found with reinforcement, probably because of the low operant rate of this response. It should be noted that the instructions contained examples of relevant responses, indicating that symbolic modeling may also have influenced the subjects' behavior. Using a more neutral response class (verbalization in a classroom setting), D'Zurilla (1966) compared the effects of persuasion (urging and prompting instructions), praise (verbal reinforcement), and a neutral interaction condition with a no-intervention control procedure. All treatment conditions were superior to the control, but did not differ significantly among each other.

Studies have also been conducted that compare the effects of instructions and reinforcement alone with a combination of the two procedures. Ayllon and Azrin (1964), in an attempt to modify the ward behavior of hospitalized psychotics, found that a combination of instructions followed by reinforcement was the most effective treatment. Reinforcement alone was ineffective (low base-rate behavior), and instructions alone produced intermediate initial effects but did not maintain the behavior over time as compared to the combination condition. Similar results were obtained by Levy (1967b), using size judgments as the dependent variable. Instructions followed by reinforcement produced the greatest effect. Reinforcement alone was not effective; intermediate effects were obtained for instructions alone. Levy accounted for the effectiveness of the combination condition by suggesting that reinforcement may serve a feedback function, confirming performance criteria specified in the initial instructions.

*Instructions versus models.* Comparisons have also been made between instructions and the use of models in the modification of verbal behavior. Jacobson[3] compared the relative effects of instructing subjects to discuss personal problems with exposure to a taped model engaged in problem disclosure. In the interview task, the results showed that instructions produced an initial facilitatory effect early in the interview session. Whalen (1969) also compared the effects of instructions and modeling upon self-disclosure in small leaderless discussion groups. Subjects were assigned to one of four conditions: filmed model of a group engaging in self-disclosure and interpersonal feedback, detailed instructions to engage in this behavior, detailed instructions plus model, and minimal instructions only (control). Only the combination of instructions and model was effective in increasing self-disclosure.

The instructions used by Whalen in the combination condition contained more than just a detailed description (defining the response class) of the target behavior. They also instructed subjects to "try and interact in a manner similar to that of the film group," and described the film as providing "excellent examples of desirable types of group behavior [p. 511]." Similar exhortations were not given in the model-only condition. This information as to the purpose of the experiment may have contributed to the superior performance of the combination condition. In the pure modeling technique, as described in this paper, instructions to engage in imitative behavior are not provided.

Instructions that contain examples of the target response, or that provide instigation to imitate the behavior of a model may confound sources of information and incentive. A recent study by Green and Marlatt[4] attempted to control for some of these factors. Subjects were given either instructions to talk about their subjective feelings, instructions to talk about their objective ideas, or received no instructions. The instructions defined the response class in detail, but provided no examples of the target behavior. Half the subjects were also exposed to a verbal model who described his feelings within specified topic areas. No instructions to imitate were provided in the modeling conditions. Separate analyses of feeling and ideational statements were conducted for verbal performance in a short self-disclosure monologue. The results for feeling statements showed that both the modeling and instructional factors were capable of significantly increasing this response, but that the instructions condition was more effective compared to modeling alone. A combination of feeling instructions and the model had the most facilitative effect. In the analyses for ideational statements, instructions and modeling were equally effective.

---

[3] E. A. Jacobson. A comparison of the effect of instructions and models upon interview behavior of high-dependent and low-dependent subjects. Unpublished doctoral dissertation, Indiana University, 1968.

[4] A. H. Green & G. A. Marlatt. The effects of instructions and modeling upon affective and descriptive self-references. Paper presented at the annual meeting of the Western Psychological Association, San Francisco, April 1971.

*Models versus verbal conditioning*. Modeling and verbal conditioning have also received attention in comparison studies; some of these comparisons have been reviewed by Kanfer (1965). In one illustrative experiment (Kanfer & Marston, 1963), the role of modeling and reinforcement was compared in a verbal conditioning task. The main comparisons involved presence of a model (with or without vicarious reinforcement) and the presence or absence of subsequent verbal reinforcement to the observer. It was found that the addition of vicarious reinforcement to the model led to significantly greater conditioning effects compared to all other groups, even in the absence of direct verbal reinforcement to the observer. Direct verbal reinforcement alone was less effective than either vicarious reinforcement alone or the combination of both reinforcement procedures. Similar results were obtained by Bandura and McDonald (1963) in a study investigating the modification of children's moral judgments.

The superior effects of vicarious reinforcement have also been demonstrated in a study by Marlatt (1970). Subjects were assigned either to a model (receiving positive, negative, or neutral vicarious reinforcement) or to a no-model control condition prior to performance in an unstructured interview. The target response was the admission of personal problems. In the interview, subjects were again randomly assigned to one of three direct reinforcement conditions (also positive, negative, or neutral). Although both vicarious and direct reinforcement factors were effective determinants of problem disclosure, the vicarious condition showed the superior effect. In a generalization interview held a week later (no reinforcement applied), only the vicarious reinforcement condition still remained a significant determinant of task performance. Wilder (1968) has also compared the relative effectiveness of modeling and reinforcement procedures for the admission of affective self-references in a quasi-interview task.

*Multiple comparisons*. Finally, two studies are described that have compared all three modification techniques with the same target response. In a study investigating original verbal behavior (uncommon word associations), Levy (1968) compared verbal reinforcement, a symbolic modeling procedure, instructions, and a combination of instructions and reinforcement. The results showed that all treatment conditions produced an increase in the target response, compared to control subjects, but that the combination of instructions and reinforcement was the most effective condition. Reinforcement only was more effective than the role model condition. Levy interpreted this latter finding by suggesting that the role model supported a variety of performance hypotheses, but that without the additional confirmation provided by reinforcement, subjects were unable to determine the correct hypothesis.

In a similar design, Masters and Branch (1969) attempted to increase the frequency of syntagmatic word associations (versus the prepotent paradigmatic response tendency in adult subjects). The authors compared the effectiveness of instructions describing the type of association wanted, with a symbolic model (presentation of examples of the target response) and verbal reinforcement for the correct response. All treatment conditions produced results significantly greater than control groups; instructions were the most effective condition, followed by the model and reinforcement groups (all significantly different from each other). In a generalization task, the instructions group did not show the same degree of transfer as did the modeling and reinforcement groups, however.

How may all these results be interpreted? Verbal reinforcement alone seems to be particularly effective when the task itself is relatively structured (as in the Taffel task). The structure of the task may provide enough information to the subject to convey awareness of the experimental purpose; reinforcement may provide a confirmatory function in this regard. The nature of the target response is also important, as several verbal conditioning studies have noted the difficulty in modifying responses that may be subject to greater inhibition (e.g., negative self-references or affect words). Prior information as to the experimental purpose, in the form of instructions or models, appears to enhance the effectiveness of verbal reinforcement as a source of confirmation. Inhibited or low base-rate response classes seem more subject to modification through instructional and modeling procedures. The addition of vicarious reinforcement to the modeling process offers additional advantages in that it provides the observer with an overview of the response and response-consequences sequence (which may only be implied in instructions).

Finally, there is the suggestion that instructions alone fail to produce general maintenance or generalization effects upon behavior change. Instructions combined with subsequent reinforcement or modeling with the addition of vicarious reinforcement appear to produce effects that are longer lasting and that generalize across similar tasks. This effect would seem to be puzzling, as instructions provide the most explicit information as to the desired task performance. Second, compared to the modeling procedure (in which subjects must first inductively arrive at a set of general rules or self-instructions prior to generating responses consistent with these rules), instructions would seem to be the simplest procedure to produce changes—rules are provided to the subject by the experimenter, thus saving the subject the trouble of figuring them out for himself. Herein lies a speculation: Behavior change may be facilitated in those cases in which the subject arrives at his own rules that determine his behavior (i.e., by ascertaining the purpose of the experiment by observing models or by receiving reinforcement feedback for his own performance). It may be that subjects who attribute changes in behavior to themselves (as opposed to attributing changes to an external source, such as the experimenter's instructions) are more likely to persist in these changes over time and across situations (cf. Davison & Valins, 1969). For lack of a better name, this might be called the "I'd rather do it myself" phenomenon.

### IMPLICATIONS FOR PSYCHOTHERAPY

The modification of verbal behavior in controlled experimental settings has little meaning unless it can be shown how such influences operate or can be applied to the nonlaboratory setting. If changes in a subject's verbal behavior in an experimental social influence setting are determined to a large degree by variables specific to that setting, then the uncritical application of these procedures to other situations is seriously in question. Psychotherapy is one area of application that should be examined in this regard (Goldstein, Heller, & Sechrest, 1966; Heller, 1970; Heller & Marlatt, 1969).

Task structure, as discussed earlier in this paper, varies as a function of information provided to the subject affecting his awareness of the experimental purpose. What is the nature of this relationship in the psychotherapy interaction? In order to claim similarity between the experimental and clinical settings in this regard, it would be necessary to ascertain the sources of information and how they affect the awareness of patients in psychotherapy. Is it that the patient attempts to discover the *therapist's* purpose, and tries to act accordingly (by attempting to please the therapist, or act in accordance with his perceived purpose), or is it that the *patient* uses the therapeutic interaction as a means of discovering information about himself?

Self-awareness in the patient, derived from the insights gained in psychotherapy, is often stated as one of the therapist's main goals. It may be the discovery of this purpose by the patient, perhaps through the interpersonal relationship that he develops with the therapist ("transference"), which is largely responsible for changes that might occur. The patient becomes aware that the therapist's purpose is to help the patient become more aware of himself. In the experimental setting, on the other hand, the subject may become aware of the experimenter's purpose that the subject perform in a manner congruent with the aims of the experiment. Unlike therapy, self-awareness in the experimental subject is *not* the aim; in fact, many experimenters attempt to control or suppress the occurrence of awareness during task performance. It may be that the locus of awareness (discovery of the experimenter's purpose, an external source; or discovery of self-awareness, an internal source) affects the stability, maintenance, and generalizability of changes that occur in either setting.

More information is needed to clarify this question of task similarity and its relation to the generalizability of behavior change. Studies investigating the demand characteristics of the therapeutic interview would be of value for this purpose. Investigation of the use and effectiveness of structuring procedures that are provided prior to a patient's involvement in therapy would also yield valuable information. The use of prestructuring devices, such as role-induction interviews (Hoehn-Saric, Frank, Imber, Nash, Stone, & Battle, 1965), pretraining tape recordings of "ideal" therapeutic behavior (Truax & Carkhuff, 1965), and written preparation materials (e.g., Duke & Frankel, 1971) have been discussed by only a few individuals. The relationship of awareness and incentive, as a function of such sources of information, to changes in behavior in naturalistic settings remains to be fully explored.

## REFERENCES

Adams, J. K. Laboratory studies of behavior without awareness. *Psychological Bulletin*, 1957, **54**, 393–405.

Alpert, J. T., & Nakamura, C. Y. Interactive effects of information and affective components of social reinforcers on independent and dependent children. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, 3, 345–346. (Summary)

Ayllon, T., & Azrin, N. H. Reinforcement and instructions with mental patients. *Journal of the Experimental Analysis of Behavior*, 1964, **7**, 327–331.

Bandura, A. Behavioral modification through modeling procedures. In L. Krasner & L. P. Ullmann (Eds.), *Research in behavior modification*. New York: Holt, Rinehart & Winston, 1965.

Bandura, A. *Principles of behavior modification*. New York: Holt, Rinehart & Winston, 1969.

Bandura, A., & McDonald, F. J. Influence of social reinforcement and the behavior of models in shaping children's moral judgments. *Journal of Abnormal and Social Psychology*, 1963, **67**, 274–281.

Cairns, R. B. The informational properties of verbal and nonverbal events. *Journal of Personality and Social Psychology*, 1967, **5**, 353–357.

Cialdini, R. B., & Insko, C. A. Attitudinal verbal reinforcements as a function of informational consistency: A further test of the two-factor theory. *Journal of Personality and Social Psychology*, 1969, **12**, 342–350.

Das, J. P. *Verbal conditioning and behaviour*. London: Pergamon Press, 1969.

Davison, G. C., & Valins, S. Maintenance of self-attributed and drug-attributed behavior change. *Journal of Personality and Social Psychology*, 1969, **11**, 25–33.

DeNike, L. D. The temporal relationship between awareness and performance in verbal conditioning. *Journal of Experimental Psychology*, 1964, **68**, 521–529.

DiVesta, F. J., & Walls, R. T. Response selection as a function of instructions and motivation under nonreinforcement conditions. *Journal of Experimental Psychology*, 1967, **73**, 365–373.

Dixon, P. W., & Oakes, W. F. Effect of intertrial activity in the relationship between awareness and verbal operant conditioning. *Journal of Experimental Psychology*, 1965, **69**, 152–157.

Duke, M. P., & Frankel, A. S. *Inside psychotherapy*. Chicago: Markham, 1971.

Dulany, D. E. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. In C. W. Ericksen (Ed.), *Behavior and awareness*. Durham, N. C.: Duke University Press, 1962.

Dulany, D. E. Awareness, rules, and propositional control: A confrontation with S-R behavior theory. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

D'Zurilla, T. J. Persuasion and praise as techniques for modifying verbal behavior in a "real-life" group setting. *Journal of Abnormal Psychology*, 1966, **71**, 369–376.

Eriksen, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review*, 1960, **67**, 279–300.

Eriksen, C. W. (Ed.) *Behavior and awareness: A symposium of research and interpretation*. Durham, N. C.: Duke University Press, 1962.

FARBER, I. E. The things people say to themselves. *American Psychologist,* 1963, **18**, 185–197.

FINLEY, J. R., & STAATS, A. W. Evaluative meaning words as reinforcing stimuli. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 193–197.

FLANDERS, J. P. A review of research on imitative learning. *Psychological Bulletin,* 1968, **69**, 316–337.

GAGNE, R. M. Problem solving. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964.

GANZER, V. J., & SARASON, I. G. Interrelationships among hostility, experimental conditions, and verbal behavior. *Journal of Abnormal and Social Psychology,* 1964, **68**, 79–84.

GOLDSTEIN, A. P., HELLER, K., & SECHREST, L. B. *Psychotherapy and the psychology of behavior change.* New York: Wiley, 1966.

GRANT, D. A. Classical and operant conditioning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964.

HAMILTON, D. L., THOMPSON, J. J., & WHITE, A. M. Role of awareness and intentions in observational learning. *Journal of Personality and Social Psychology,* 1970, **16**, 689–694.

HEKMAT, H., & LEE, Y. B. Conditioning of affective self-references as a function of semantic meaning of verbal reinforcers. *Journal of Abnormal Psychology,* 1970, **76**, 427–433.

HELLER, K. Laboratory interview research as analogue to treatment. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change.* New York: Wiley, 1970.

HELLER, K., & MARLATT, G. A. Verbal conditioning, behavior therapy and behavior change: Some problems in extrapolation. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status.* New York: McGraw Hill, 1969.

HOEHN-SARIC, R., FRANK, J. D., IMBER, S. D., NASH, E. H., STONE, A. R., & BATTLE, C. C. Systematic preparation of patients for psychotherapy: I. Effects on therapy behavior and outcome. *Journal of Psychiatric Research,* 1965, **2**, 267–281.

HOLMES, D. S. Verbal conditioning, or problem-solving and cooperation? *Journal of Experimental Research in Personality,* 1967, **2**, 289–294.

HUNT, E. B. *Concept learning: An information processing approach.* New York: Wiley, 1962.

INSKO, C. A., & CIALDINI, R. B. A test of three interpretations of attitudinal verbal reinforcement. *Journal of Personality and Social Psychology,* 1969, **12**, 333–341.

KAHN, E. J. Verbal conditioning of positive and negative self-reference statements under three types of experimenter intervention. *Proceedings of the 75th Annual Convention of the American Psychological Association,* 1967, **2**, 257–258. (Summary)

KANFER, F. H. Vicarious human reinforcement: A glimpse into the black box. In L. Krasner & L. P. Ullmann (Eds.), *Research in behavior modification.* New York: Holt, Rinehart & Winston, 1965.

KANFER, F. H. Verbal conditioning: A review of its current status. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general S-R theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

KANFER, F. H., & MARSTON, A. R. Verbal conditioning, ambiguity, and psychotherapy. *Psychological Reports,* 1961, **9**, 461–475.

KANFER, F. H., & MARSTON, A. R. Human reinforcement: Vicarious and direct. *Journal of Experimental Psychology,* 1963, **65**, 292–296.

KANFER, F. H., & MCBREARTY, J. F. Verbal conditioning: Discrimination and awareness. *Journal of Psychology,* 1961, **52**, 115–124.

KENNEDY, T. D. Verbal conditioning without awareness: The use of programmed reinforcement and recurring assessment of awareness. *Journal of Experimental Psychology,* 1970, **84**, 487–494.

KESSEL, P., & BARBER, K. J. Experimenter-subject interaction in verbal conditioning: Review of the literature. *Psychological Reports,* 1968, **22**, 59–74.

KRASNER, L. The therapist as a social reinforcement machine. In H. H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy.* Vol. 2. Washington, D. C.: American Psychological Association, 1962.

KRASNER, L. Verbal conditioning and psychotherapy. In L. Krasner & L. P. Ullmann (Eds.), *Research in behavior modification.* New York: Holt, Rinehart & Winston, 1965.

KRASNER, L., & ULLMANN, L. P. Variables affecting report of awareness in verbal conditioning. *Journal of Psychology,* 1963, **56**, 193–202.

LEVY, L. H. Awareness, learning and the beneficient subject as expert witness. *Journal of Personality and Social Psychology,* 1967, **6**, 365–370. (a)

LEVY, L. H. The effects of verbal reinforcement and instructions on the attainment of size constancy. *Canadian Journal of Psychology,* 1967, **21**, 81–91. (b)

LEVY, L. H. Originality as role-defined behavior. *Journal of Personality and Social Psychology,* 1968, **9**, 72–78.

MANDLER, G., & KAPLAN, W. K. Subjective evaluation and reinforcing effect of verbal stimulus. *Science,* 1956, **124**, 582–583.

MARLATT, G. A. Comparison of vicarious and direct reinforcement control of verbal behavior in an interview setting. *Journal of Personality and Social Psychology,* 1970, **16**, 695–703.

MARLATT, G. A. Exposure to a model and task ambiguity as determinants of verbal behavior in an interview. *Journal of Consulting and Clinical Psychology,* 1971, **36**, 268–276.

MARLATT, G. A., JACOBSON, E. A., JOHNSON, D. L., & MORRICE, D. J. Effect of exposure to a model receiving evaluative feedback upon subsequent behavior in an interview. *Journal of Consulting and Clinical Psychology,* 1970, **34**, 104–112.

MARX, M. H. Some suggestions of the conceptual and theoretical analysis of complex intervening variables in problem-solving behavior. *Journal of General Psychology,* 1958, **58**, 115–128.

MASTERS, J. C., & BRANCH, M. N. Comparison of the relative effectiveness of instructions, modeling, and reinforcement procedures for inducing behavior change. *Journal of Experimental Psychology,* 1969, **80**, 364–368.

Matarazzo, J. D., Saslow, G., & Pareis, E. N. Verbal conditioning of two response classes: Some methodological considerations. *Journal of Abnormal and Social Psychology,* 1960, **61**, 190–206.

Merbaum, M. The conditioning of affective self-references by three classes of generalized reinforcers. *Journal of Personality,* 1963, **31**, 179–191.

Merbaum, M., & Lukens, H. C. Effects of instructions, elicitations, and reinforcements in the manipulation of affective verbal behavior. *Journal of Abnormal Psychology,* 1968, **73**, 376–380.

Oakes, W. F. Verbal operant conditioning, intertrial activity, awareness, and the extended interview. *Journal of Personality and Social Psychology,* 1967, **6**, 198–202.

Postman, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964.

Postman, L., & Sassenrath, J. The automatic action of verbal rewards and punishments. *Journal of General Psychology,* 1961, **65**, 109–136.

Sarason, I. G. Experimental approaches to test anxiety: Attention and the uses of information. In C. D. Spielberger (Ed.), *Anxiety and behavior.* Vol. 2. New York: Academic Press, in press.

Simkins, L. Instructions as discriminative stimuli in verbal conditioning and awareness. *Journal of Abnormal and Social Psychology,* 1963, **66**, 213–219.

Skinner, B. F. *Verbal behavior.* New York: Appleton-Century-Crofts, 1957.

Spielberger, C. D. The role of awareness in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N. C.: Duke University Press, 1962.

Spielberger, C. D. Theoretical and epistemological issues in verbal conditioning. In S. Rosenberg (Ed.), *Directions in psycholinguistics.* New York: Macmillan, 1965.

Spielberger, C. D., Bernstein, I. H., & Ratliff, R. G. The information and incentive value of the reinforcing stimulus in verbal conditioning. *Journal of Experimental Psychology,* 1966, **71**, 26–31.

Spielberger, C. D., & DeNike, L. D. Descriptive behaviorism versus cognitive theory in verbal operant conditioning. *Psychological Review,* 1966, **73**, 306–326.

Spielberger, C. D., & Levin, S. M. What is learned in verbal conditioning? *Journal of Verbal Learning and Verbal Behavior,* 1962, **1**, 125–132.

Spielberger, C. D., Levin, S. M., & Shepard, M. C. The effects of awareness and attitude toward the reinforcement on the operant conditioning of verbal behavior. *Journal of Personality,* 1962, **30**, 106–121.

Taffel, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology,* 1955, **51**, 496–501.

Tatz, S. J. Symbolic activity in "learning without awareness." *American Journal of Psychology,* 1960, **73**, 239–247.

Thorndike, E. L. *The psychology of wants, interests, and attitudes.* New York: Appleton-Century, 1935.

Truax, C. B., & Carkhuff, R. R. Personality change in hospitalized mental patients during group psychotherapy as the function of the use of alternate sessions and vicarious therapy pretraining. *Journal of Clinical Psychology,* 1965, **21**, 225–228.

Verplanck, W. S. Unaware of where's awareness: Some verbal operants—notates, monents, and notants. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N. C.: Duke University Press, 1962.

Weiss, R. L., Ekman, P., Ullmann, L. P., & Krasner, L. The context of reinforcement in verbal conditioning. *Journal of Clinical Psychology,* 1965, **21**, 99–100.

Whalen, C. Effects of a model and instructions on group verbal behaviors. *Journal of Consulting and Clinical Psychology,* 1969, **33**, 509–521.

Wilder, S. M. The effect of verbal modeling and verbal reinforcement on the frequency of self-referred affect statements. *Dissertation Abstracts,* 1968, **28**, 4304B–4305B.

Zegers, R. A. Expectancy and the effects of confirmation and disconfirmation. *Journal of Personality and Social Psychology,* 1968, **9**, 67–71.

(Received April 5, 1971)

Psychological Bulletin
1972, Vol. 78, No. 5, 351–352

# GENE ACTION AND BEHAVIOR OF C3H AND DBA MICE

## A REPLY TO WILCOCK AND THIESSEN

DANIEL BOVET,[1] FILOMENA BOVET-NITTI, AND ALBERTO OLIVERIO

*Laboratorio di Psicobiologia e Psicofarmacologia, Rome*

In commenting on a review by Wilcock, the problem of retinal degeneration (*rd*) of C3H mice is raised. In particular, the performance of C3H mice is discussed in relation to that of other strains in terms of *rd*. It is pointed out that C3H mice achieve high levels of responding under massed practice when a visual conditioned stimulus is adopted and that comparable levels of avoidance are reached when stimuli other than light are adopted.

A thorough review has been published by Wilcock (1969) on single-gene studies. In his comments to this review, Thiessen (1971) pointed out that a physiological analysis of single-gene substitution may provide simpler descriptions of otherwise obscure phenotypes. In this regard, a recent paper by Bovet, Bovet-Nitti, and Oliverio was examined with particular emphasis on the shuttle box performance of the C3H strain using the outset of light as a conditioned stimulus. The inferiority of C3H mice in relation to other strains is related by Thiessen to the fact that many C3H strains suffer from retinal degeneration (*rd*). In answer to Thiessen, Wilcock (1971) concluded that it is impossible to call these mice "cognitively deficient," and that the explanation of the results of Bovet, Bovet-Nitti, and Oliverio (1969a, 1969b) is trivial in light of the "widely published and easily accessible" findings on the retinal degeneration of C3H mice (Sidman & Green, 1965).

In answer to a similar suggestion (Henry, Buckholtz, & Bowman, 1969), it was clearly pointed out (Bovet et al., 1969a, 1969b) that (*a*) comparable levels of avoidance responding were elicited when auditory stimuli rather than lights were used; (*b*) comparable levels of transfer were attained when C3H mice were shifted from light to tone as from tone to light (Oliverio & Bovet, 1969); (*c*) C3H mice reached above-chance levels in a fire-choice discrimination task when sets of patterns differing by their spatial orientation were adopted (Bovet-Nitti, 1969).

Further, it is necessary to point out that the main point of the research on the avoidance performance of different strains was not concerned with a classification of the strains on the basis of their "cognitive abilities." The C3H mice were shown to be inferior to the DBA strain when the trials were spaced, but high levels of performance were instead reached under massed practice. It must be stressed that the main point of the research consisted in the demonstration that opposite performance levels were reached by the two strains depending on the schedule adopted (massed versus distributed practice) and independently on the type of stimulus (visual or auditive) employed. Thus, the fact that C3H/He mice attain high performance level under massed practice or when an auditive conditioned stimulus was adopted was clearly evident and in contrast to an interpretation of the results on the ground of the *rd* hypothesis.

Since then other findings dealing with the visual perception of C3H mice have been published and are easily accessible.

Nagy and Forrest (1970) have tested the open-field behavior of C3H mice under conditions of different illumination levels. The overall activity was shown to decrease as the illumination levels increased. In another research the ability of C3H to respond to visual stimuli was examined at different ages (Nagy & Misanin, 1970).

The pattern perception on the visual-cliff task showed that C3H mice are capable of pattern discrimination through 40 days of age, while a lower ability is evident at older ages. In a second experiment it was shown that these mice are capable of brightness discrimination after 40 days of age. In addition to that, it has previously been shown that *rd* mice are capable of brightness discrimination when sufficient levels of illumination are employed (Karli, 1952). Pattern discrimination was also demonstrated in *rd* mice (Karli, 1954). In agreement with Thiessen (1971), the findings quoted above demonstrate that the many single-gene studies of behavior concentrate also on central phenomena and not exclusively on empirical or "trivial" findings dealing with obvious peripheral mechanisms (Wilcock, 1971).

---

[1] Requests for reprints should be sent to Daniel Bovet, Laboratorio di Psicobiologia e Psicofarmacologia, del C. N. R., Via Reno 1. 00198 Roma.

## REFERENCES

Bovet, D., Bovet-Nitti, F., & Oliverio, A. Genetic aspects of learning and memory in mice. *Science,* 1969, **163**, 139–149. (a)

Bovet, D., Bovet-Nitti, F., & Oliverio, A. Genetics of Memory. *Science,* 1969, **165**, 1148. (b)

Bovet-Nitti, F. Facilitation of simultaneous visual discrimination by nicotine in four inbred strains of mice. *Psychopharmacologia,* 1969, **14**, 193–199.

Henry, K. R., Buckholtz, N. S., & Bowman, R. E. Genetics of memory. *Science,* 1969, **165**, 1148.

Karli, P. Rétines sans cellules visuelles-recherches morphologiques, physiologiques, et physiopathologiques chez les rongeurs. *Archives d'Anatomie Histologie et Embryologie,* 1952, **35**, 1–76.

Karli, P. Etude de la valeur fonctionelle d'une rétine dépourvue de cellules visuelles photo-receptices. *Archives de Sciences Physiologiques,* 1954, **8**, 305–328.

Nagy, Z. M., & Forrest, E. J. Open-field behavior of C3H mice: Effect of size and illumination of field. *Psychonomic Science,* 1970, **20**, 19–20.

Nagy, Z. M., & Misanin, J. R. Visual perception in the retinal degenerate C3H mouse. *Journal of Comparative and Physiological Psychology,* 1970, **72**, 306–310.

Oliverio, A., & Bovet, D. Transfer of avoidance responding between visual and auditive stimuli presented in different temporal patterns. *Communications in Behavioral Biology,* 1969, **3**, 61–68.

Sidman, R. L., & Green, M. C. Retinal degeneration in the mouse: Location of the *rd* locus in the linkage group XVII. *Journal of Heredity,* 1965, **56**, 23–29.

Thiessen, D. D. Reply to Wilcock on gene action and behavior. *Psychological Bulletin,* 1971, **75**, 103–105.

Wilcock, J. Gene action and behavior: An evaluation of major gene pleiotropism. *Psychological Bulletin,* 1969, **72**, 1–29.

Wilcock, J. Gene action and behavior: A clarification. *Psychological Bulletin,* 1971, **75**, 106–108.

(Received April 6, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 5, 353–367

# AFRICAN INFANT PRECOCITY

NEIL WARREN[1]

*University of Sussex, Brighton, England*

Studies of African infant development are reviewed with reference to the possible phenomenon of African infant precocity. Although the majority of studies report precocity, it is held that defects of measurement and design must preclude the conclusion that precocity is an established fact. In particular, the conclusion of precocity has always been based on comparison with test norms, whereas studies involving actual comparative samples do not find precocity. It is also argued that infant differences by social milieu afford the most sensible basis for the necessary introduction of independent variables into this research area, and that improved techniques of assessment should be applied, both in the neonatal period and beyond.

Research on the behavioral development of the African infant is reviewed, with special reference to the putative phenomenon of African precocity in the first year or more of life. Although Africa embraces a number of racial types or subtypes and great cultural diversity, as a first step studies are brought together from East, West, and South Africa, plus Madagascar. The term "African" here refers to the black peoples of sub-Saharan Africa.

Research effort to date has gone little further than mere assessment of the developmental levels of some available infants. If African precocity be a genuine phenomenon, at least for some groups, and can be accounted for by well-designed comparative studies, answers to general questions about the effects of early stimulation and experience and, conceivably, to questions of behavioral genetics may be forthcoming. Additional, broader developmental questions would include the relation between infant development and development beyond infancy. Cross-cultural study should be able to take advantage of a greater range and diversity of independent and dependent variables than Western culture offers. At its most ambitious, the rationale of the cross-cultural enterprise is to facilitate the open-mindedness recommended by Feyerabend (Kessel, 1969), that of actively seeking to invent alternatives to existing theory and assumption. Appropriately, some of the findings reviewed are at variance with predictions that would be made from assumptions concerning infant development in the West. At the more conventional level of hypothesis-testing research, African infant precocity research falls primarily within the strategy that Levine (1970) suggested is preferable in cross-cultural work—that of beginning with an identified phenomenon as a dependent variable. This paper assesses the documentation of a putative phenomenon which may, if it exists, be explained.

A further aim is to make available an account, in English, of reports that, for reasons of diversity and obscurity, are unlikely to be encountered. Citations (often erroneous) from a single brief and inadequate paper (Geber, 1958a) have usually been the basis for generalizations on African infant development. The full range of reports includes many in French, some from medical or Africanist journals, and a number of degree dissertations.

## INFANT AND EARLY CHILD DEVELOPMENT (EXCLUDING NEWBORNS)

A brief summary of the essential details of each study is found in Table 1. Only one pre-World-War-II report (Langton, 1934–1935) deals at all quantitatively with African infant development. Langton recorded ability to crawl, stand, and walk for 76 infants in Bunyoro, Uganda, concluding, without explicit comparisons with British data, that the Munyoro infant is not more precocious than the British infant.

---

[1] Requests for reprints should be sent to Neil Warren, School of African and Asian Studies, University of Sussex, Brighton, BN1 9QN, England.

TABLE 1

SUMMARY OF STUDIES

| Author(s) and date | Country | *N* |  | Age range | Test used | Reported findings | Comments |
|---|---|---|---|---|---|---|---|
|  |  | Africa | Europe |  |  |  |  |
| Langton (1934–1935) | Uganda | 76 | 0 | 6½–17 months | Observation in clinic | Africans not more precocious than British | No explicit comparison made. Sketchy report. |
| Theunissen[a] | South Africa | 66 | 66 | 2–57 weeks | Gesell | No general differences. | Very detailed and careful report. |
| Faladé (1955) | Senegal | 101 | 0 | 4–56 weeks | Gesell | Consistant precocity against Gesell norms. | Longitudinal testing analyzed cross-sectionally. |
| Geber (1956, 1958a, 1958b, 1960, 1962); Geber & Dean (1957a, 1957b, 1958, 1964) | Uganda (also Senegal and South Africa) | 252 (357 testings) | 0 | 1 month–6 years | Gesell; Terman-Merrill | Precocity against norms, declining with age. Differences by social level among Africans. | Some inconsistencies between articles. Incomplete reporting of data. Longitudinal testing analyzed cross-sectionally. |
| Falmagne ([1959],[b] 1962) | South Africa | 78 | 105 | 1–6 months | Own items and indexes | No general differences. See text. | Very detailed and careful report. |
| Vouilloux (1959) | Cameroons | 48 | 0 | 1–36 months | Gesell | Precocity against norms, declining with age. Differences by social level. | No comment. |
| Ramarosaona[c] | Madagascar | 48 | 0 | Up to 5 years | Gesell-based observation | Early precocity until "pause in development." Difference by social level. | Very sketchy presentation. Unconvincing documentation of precocity. |
| Massé (1969) | Senegal | 1,162 testings | 0 | 6 months–6 years | Brunet-Lézine; Terman-Merrill | Precocity in first year declining to retardation after age 2. | Very detailed presentation. Longitudinal testing analyzed cross-sectionally. |
| Ainsworth (1963, 1967) | Uganda | Up to 28 | 0 | 15–35 weeks | Observation in home | Precocity in motor development and language, against test norms. | Incidental aspect of study of attachment. |
| Liddicoat & Koza (1963) | South Africa | 120 | 0 | 18–36 months | Own items; Gesell-based | Language development up to and surpassing Gesell norms | Items selected for cultural suitability |
| Liddicoat (1969) | South Africa | 480 | 0 | 1–12 months | 6 motor items | Precocity against all Western test norms for these items. | Preliminary report of study to develop African baby test. |
| Kilbride[d] | Uganda | 163 | 0 | 1–24 months | Bayley Motor Scale | Consistent precocity against norms declining with age. | Precocity much greater by earlier version of Bayley test. |
| Goldberg[e] | Zambia | 38 | 0 | 4 months | Bayley Motor Scale | Precocity against norms. | No correlations between DQ and measures of maternal stimulation. |
| Lusk & Lewis (in press) | Senegal | 10 | 0 | 2–12 months | Bayley Motor & Mental Scales | Precocity against norms—greater for motor development. | No correlations between DQ and measures of mother–infant interaction. |

[a] See Footnote 2, text.
[b] See Footnote 4, text.
[c] See Footnote 5, text.
[d] See Footnote 6, text.
[e] See Footnote 7, text.

Geber, often cited as pioneer, was preceded by Theunissen[2] and Faladé (1955). Theunissen tested 232 African and European[3] infants in two urban centers and one village in South Africa. For comparative purposes, 66 European and 66 African infants whose ages corresponded to the nearest week were selected. African families were described as urbanized and semi-urbanized; Europeans as "fairly middle-class." Theunissen provided charts for each item at each key age level, and it is appropriate to quote his own conclusion:

---

[2] K. B. Theunissen. A preliminary comparative study of the development of motor behavior in European and Bantu children up to the age of one year. Unpublished master's thesis, Natal University College, Durban, South Africa, 1948.

[3] The term "European" means here "of European stock," that is, white. It is used below in this sense, which is the established convention throughout the African continent.

> a close examination of the charts reveals a marked degree of similarity in the rate of development of European and Bantu infants. In no schedule does the aggregate picture show any superiority for either race, and if any item is followed in its progressions through different schedules, the same conclusion is reached [p. 83].

Theunissen viewed this as supporting the belief that the African is not genetically inferior to the European, in the first year of life.

Faladé (1955) tested 101 healthy Senegalese children: 75 urban children were tested each month for 5 months: and 26 rural children were tested once only. No distinction was made between these groups in the report, and findings were presented cross-sectionally. Senegalese precocity over the Gesell's (Gesell Developmental Schedules) normative base line was found at every age level from 4 to 56 weeks and for every aspect of development measured, the greatest differences appearing at 36–40 weeks of age. Locomotor development showed the greatest advance up to 36 weeks, but was subsequently overtaken by other behaviors. Faladé noted the development of language behavior, instancing children of 11–12 months who possessed a vocabulary of a dozen or so words. She declined to present quotients, because of the absurdly high quotients a small chronological advance can give at early ages. By the reviewer's calculation, mean developmental quotients (DQs) were 163 at 4 weeks, 132 at 28 weeks, and 114 at 56 weeks, representing mean chronological advances of about 2½ weeks, 9 weeks, and 8 weeks, respectively. A further paper by Faladé (1960) qualitatively summarizes Senegalese infant development to a point well beyond the first year, but without presentation of additional data.

One sketchy article (mostly photographs), which appeared in English in a psychological journal (Geber, 1958a), is the paper by which Geber's work tends to be known among English-speaking psychologists. A summary, which is culled from the relevant publications from 1956 to 1964, of Geber's work is given. With two minor exceptions, Geber has worked in Uganda. Geber (1960, p. 108) performed 357 assessments of children aged from 1 month to 6 years: this involved 252 subjects in all, since unspecified numbers were tested twice or thrice, and, in the 1–36-month range, 224 testings were conducted. No longitudinal analysis was made, and reporting was in fact casual, often little more than anecdotal. There are also some inconsistencies from article to article. The most complete quantitative presentation (Geber, 1960) merely gave frequencies and percentages at each age level falling within given ranges of Gesell DQ points, including the hundred or so retests as independent assessments. These results indicate a remarkable precocity in the first year (100% ≧ 100 DQ), which declines steadily to a low point in the third year (54% ≧ 100 DQ), but which is then regained and maintained up to the sixth year (88% ≧ 100 DQ). Whatever this precocity, it was not typically of the magnitude to which a number of American authors (e.g., Beadle, 1970; Casler, 1961; Thompson, 1968) have referred: that Ugandan babies sit at 7 weeks and walk at 7 months. These statements arise, it would seem, because Geber (1958a) included photographs to illustrate precocity. One of these (Geber's Figure 3) purported to show a 7-week-old "supporting herself in a sitting position." Whatever is meant by this statement, it is not clear from the photograph that the child is sitting unsupported, from the positions of the legs and the visible buttock; in the text, Geber gave 4 months as the age at which Ugandan babies sat alone. Another photograph (Geber's Figure 6) showed a 7-month-old standing at the Gesell box, purportedly having walked to it to look inside. In the text, it is asserted that 10 months was the average age for walking. The photographs in question clearly show extreme rather than typical examples.

Geber also tested 30 African infants both in South Africa and in Senegal: The results are not given, however, and she simply remarked that an equivalent precocity was found (Geber, 1960, p. 104). Because she used open-ended categories of DQ range, it is not possible to calculate from Geber's presentation of her Ugandan results what the mean DQ's or chronological advances at each age level are. She stated (Geber & Dean,

TABLE 2

Average Developmental Quotient (DQ) for Each Age Range

| Mean age (months) | Mean DQ | Mean advance (weeks) |
|---|---|---|
| 0–3 | 134.5 | 3.0 |
| 4–7 | 125.8 | 6.7 |
| 8–11 | 109.8 | 4.3 |
| 12–19 | 106.0 | 4.2 |
| 20–36 | 107.1 | 8.6 |

1964, p. 427) that an advance of 2–3 months was maintained from 3 to 18 months of age—which would mean a steadily declining mean DQ from around 180 to around 115. By subscales, precocity was described as most marked in motor development during the first few months, but in all subscales equally by 6 months: The general advance persists until about 18 months, when a "stagnation" of development sets in. Apart from categories of sex, tribe (no differences by either), age, and social level, Geber gave no empirical analysis of any independent variable in relation to development. This is surprising because (*a*) she speculated or pronounced on determinants in every article, so that at one point or another genetic factors, maternal attitudes, nutrition, weaning, and various aspects of care, handling, and rearing were implicated; (*b*) she reported throughout (e.g., Geber, 1956, 1958a, 1960) that the mothers were interviewed concerning the child's treatment and behavior, information that appears to have languished unanalyzed.

To test children between 3 and 6 years, Geber used both the Terman-Merrill Revision and the Gesell developmental norms. It can be calculated, from Geber (1960, p. 109), that 97 testings were conducted with the Terman-Merrill Revision. The Terman-Merrill is not mentioned in the well-known paper (Geber, 1958a), and Geber is clearly dissatisfied with it in this cultural context (e.g., Geber, 1958b, p. 518; 1960, p. 102; Geber & Dean, 1958; 1964, p. 435), so that she accords priority to Gesell DQs in reporting performance even up to 6 years. It is therefore curious that, despite Geber's reservations, the Terman-Merrill seems to have produced higher quotients than the Gesell (Geber & Dean, 1964, p. 433). If we take Geber's (1960) results by Gesell alone or by both tests, it seems fair to say that precocity is sustained throughout with a marked dip in the third year only. Earlier assertions by Geber (e.g., 1956, 1958a, 1958b), implying that the precocity disappeared for good after the first year or two, have been repeated in the American literature (e.g., Ainsworth, 1967; Jahoda, 1968; Levine, 1970). To confuse matters further, Geber and Dean's (1964, p. 433) graph, based on the same data, is difficult to square with the figures given by Geber (1960). The graph places the "traditional" African mean DQ below the Gesell average from 4 to 6. It is not possible to compare these presentations precisely, owing to the open-ended DQ ranges in Geber (1960).

Falmagne ([1959],[4] 1962) compared 78 African infants in Johannesburg with 105 Belgian infants in Brussels. Age range was from 1 to 6 months (when the greatest precocity against norms is reported by several other studies), and the assessment comprised 20 behavioral items or indexes of development: 11 items were measures of motor control; 9 were perceptual or sensorimotor responses to stimuli.

Falmagne provided a careful and explicit report. A comparison can be made for each month of age on each of 20 items. Omitting those comparisons on which the groups are equal, 79 comparisons remain: 42 of these favor the European babies and 37 the African babies. On the perceptual and sensorimotor items alone, 22 comparisons favor Europeans, 15 Africans. Falmagne took a larger sample of African babies for this age range than did any other investigator (except Liddicoat, 1969): for example, 78 compared with Geber's 34. The range of items was narrow compared with that of a comprehensive baby test, but Falmagne included a few items that Geber singled out as examples of remarkable precocity in African infants.

Vouilloux (1959) tested 48 "normal" infants in the Cameroons. Average DQs for each age range are given in Table 2. No numbers by age were given. The mean advance over Gesell norms for each age group has been calculated by the reviewer, taking the center of the age range as the "chronological age." From the mean advance in the highest age range, there appears to be no basis for Vouilloux's extrapolation to the undocumented statement (p. 13) that precocity disappears after age 3. It is possible that this statement reflects the already published assertions of Geber on the matter.

---

[4] J.-Ce. Falmagne. Étude de certains aspects du développement du nourrisson de 0 a 6 mois. Unpublished doctoral thesis, Université Libre de Bruxelles, Bruxelles, 1959.

The Cameroon babies were most advanced in locomotion (mean DQ = 133) and not at all on the vocal–verbal scale (mean DQ = 100). Vouilloux also included a number of comparisons between African groups: She tested 25 children, hospitalized for illness, and 25 orphanage children. Mean DQs were 117 for the "normal" group discussed above, 101 for the sick group, and 78 for the orphanage group (calculated from Vouilloux, 1959, p. 13).

Ramarosaona [5] briefly reported observations of 48 Merina children in Madagascar. They were observed "continuously for a certain length of time" in a temporary home after a flood, using the Gesell schedules as a rough basis for observations, apparently of motor and language development only. Numbers available by years of age were as follows: 0–1 year, 10; 1–2 years, 13; 2–3 years, 10; 3–5 years, 15. The author stated that for the first year the Merina infant is at the same level as Geber's Ugandan children, with an advance of about 2 months over Gesell norms. From 12 months a "pause in development" was noted. There is no data presentation, but, aspects of locomotion in the first year apart, Ramarosaona's account of verbal development does not convincingly suggest precocity. Walking and manual dexterity are not obviously advanced, and the report of language development (based on conversations with the mothers) implies, if anything, retardation (e.g., first word "mama" at 12 months.) It is not possible to evaluate Ramarosaona's conclusions more precisely, and DQs are not given or calculable. It is the reviewer's impression that Ramarosaona may have been enthusiastic to see herself as replicating Geber's findings, without observational material extensive or precise enough to do so.

Bardet, Massé, Moreigne, and Senecal (1960) and Massé (1969) reported a program of studies of Senegalese infants. The Brunet-Lézine test, a Gesell-inspired French scale, was administered to subjects of 6 months (57), 9 months (91), 12 months (144), 18 months (129), and 2 years (109), of a "modest" social background. The results are summarized (Massé, 1969, pp. 186–187, p. 234) as revealing a mean advance of 2 months over French norms between 6 and 12 months, dropping to about the French level from 12 to 18 months, with a "stagnation" of development toward 24 months. Detailed tables give percentages passing each item, but do not allow calculation of mean DQs or chronological advances. It is difficult to see evidence in these tables of a sudden stagnation: there is rather a steady decline vis-à-vis French norms from 6 to 24 months which flattens into a plateau (which could even be called a "spurt") at about the French average level between 12 and 18 months. Geber, of course, had similarly spoken of a "stagnation" of development, after 18 months. A further observation to be made from the tables is that there is a consistent backwardness on the language items, even in the first year.

Moreigne and Senecal (1962) and Massé (1969) further reported the use of a slightly modified Terman-Merrill Revision to test 66 children of 3 years, 199 of 4 years, 254 of 5 years, and 113 of 6 years, including an unspecified number of children tested twice or thrice in successive years. Mean quotients decline gradually with age, from 95.4 at age 3 (which can be judged by inspection to be close to the mean Senegalese Brunet-Lézine DQ at age 2, though this is not given or precisely calculable from Massé's ([1969] tables) to 93.2 at age 4, 88.4 at age 5, and 87.5 at age 6, girls scoring slightly higher than boys at every age. Correlations from year to year for subjects tested more than once were .63 between age 3 and 4, .76 between age 4 and 5, .78 between age 5 and 6, and .62 between age 3 and 5. These (which are taken by Massé as reliability coefficients) compare nicely with American Terman-Merrill data for these ages (e.g., Ebert & Simmons, 1943; Sontag, Baker, & Nelson, 1958), and show the usual trend of an increasing correlation with age and a decreasing one as the interval between testings is lengthened. They may be taken as some indication that the test is not unsuitable for Senegalese culture and that intelligence shows a degree of stability over this age range that is comparable with that observed in the United States. The slight superiority of girls at these age levels is also consistent with American findings (e.g., Ebert & Simmons, 1943; Terman & Merrill, 1960).

---

[5] Z. Ramarosaona. Psychomotor development in early childhood in the Tananarive region. Annex IX in Conseil Scientifique pour L'Afrique au Sud du Sahara meeting of specialists on the basic psychology of African and Madagascan populations. (Conseil Scientifique pour L'Afrique au Sud du Sahara. Publication No. 51) Tananarive, Madagascar, 1959.

Ainsworth (1963, 1967) made systematic observations of 28 Ugandan infants in their homes over periods varying from about 15 to 35 weeks. Records of certain milestones in motor development were kept, and Ainsworth reported that her babies "sat and crawled and stood and walked much earlier than the average baby in Western societies [1967, p. 319]." Crawling averaged 2½ months earlier than Griffiths' (1954) norms and creeping about 3½ months earlier (Ainsworth, 1967, p. 325). Ainsworth also recorded observations suggesting precocity of language development. Her main research emphasis was on attachment behavior, and although it is possible that Ugandan babies are precocious in this respect (e.g., separation anxiety present by 6 months), the sample size and methods prevent adequate comparisons. Kagan's (1969) assertion of a Ugandan advance of 4 months over American development of attachment to a caretaker appears to set the American age level unjustifiably high.

Liddicoat and Koza (1963) asked 120 African children, male and female, at age levels from 18 to 36 months, to name and identify picture cards or objects specially selected for their cultural suitability. On the whole, performance was well up to Gesell's norms for similar procedures, and on some comparisons surpassed them by wide margins: On the average, 62% reached Gesell's norms on picture response (notwithstanding the results of Hudson, 1967), and 59% on naming and giving the use of objects. In the third year, naming responses were particularly advanced: 70% surpassed Gesell norms for pictures, and 92% for objects (calculated from Liddicoat and Koza's tables).

Liddicoat (1969) also briefly reported the testing of a cross-sectional sample, 20 boys and 20 girls for each month of the first year, using items from various baby tests. African normative age levels were ahead, and mostly well ahead, of all Western norms on each of six motor items reported. Kilbride[6] employed Bayley's Infant Scale of Motor Development, in both its 1936 and its revised 1965 form, to cover key ages from 1 to 24 months in Ugandan infants. She provided a full and careful report of her results. Motor DQs were calculated according to both versions of the Bayley test. On the revised version, mean, median, and modal DQ were 123, 123, and 124, respectively; by the 1936 test, they were 145, 142, and 139. On the revised test, 75% scored above 110 DQ and 96% on the 1935 test. On the 1965 test, 1% were below 90 DQ and none at all on the earlier version. Differences between revised Bayley normative means and Ugandan means are highly significant at all age levels. Precocity vis-à-vis Bayley norms was greatest at 4 months of age, subsequently decreasing with age, but Ugandans were still significantly advanced at 24 months. Mean DQs for the four successive 6-month periods were 131, 121, 116, and 109 on the revised test. The substance of this report has been subsequently published by Kilbride, Robbins, and Kilbride (1970).

Two further studies were also based on the new Bayley test. Goldberg[7] used the motor scale to test 38 Zambian infants at 4 months of age. Mean DQ was 126, an advance of roughly 3 weeks over American norms, and it appears from the histogram provided that only four babies had DQs of 100 or less. Lusk and Lewis (in press) employed both the mental and motor Scales with 10 infants in Senegal whose ages varied between 2 and 12 months. Motor DQs were very high: six exceeded the scorable limit (150) of the Bayley index, and only one was below 100, just. Mental DQs were less extreme, but high: eight were around 130, or above, and only below 100. Performance represented, on average, a motor advance of about 2 months and a "mental" advance of about 5 weeks. Both Goldberg and Lusk and Lewis took observational measures of mother's and infant's behavior in time samples, but in neither study did significant correlations appear between variables of mother–infant interaction and DQ.

---

[6] J. E. Kilbride. The motor development of rural Baganda infants. Unpublished master's thesis, Pennsylvania State University, 1969.

[7] S. Goldberg. Infant care in Zambia. (Report No. 13) Human Development Research Unit, University of Zambia, 1970.

### *Is There a Phenomenon To Be Explained?*

#### *Tally*

If we count studies according to their own conclusions, there is a large majority in favor of African infant precocity. Counting Geber's work as one investigation, and omitting Langton (1934–1935) and Liddicoat and Koza (1963) as too narrowly focused, the tally is 10 studies for and 2 against.

#### *Pattern*

There is reasonable consensus that DQ precocity is greatest in the first 6 months and declines thereafter; and that motor development shows the greatest advance up to 6 or 9 months of age, but subsequently other subscales show a roughly comparable advance. The only clear disagreement is on vocal–verbal behavior, for which Vouilloux could find no advance at all and the Senegal group found retardation, whereas Faladé, Geber, and Ainsworth noted precocity, and Liddicoat and Koza, in a specifically focused and careful study, found performance as good as or better than Gesell normative levels even in the second and third year. There is also no agreement whether or when the precocity may disappear altogether. Massé's summary indicates that precocity disappears by age 2, and the Senegalese Terman-Merrill results, indicating a continuing gradual decline from age 3 to 6, have parallels with American Terman-Merrill research. It remains, however, that the Terman-Merrill is probably more culturally biased than a baby test, and increasingly so with increasing age. Geber seems to contradict herself from one presentation of findings to the next, but her fullest display of results undoubtedly indicates that precocity is found up to age 6. If we also take into account Liddicoat and Koza's (1963) findings of excellent verbal performance in the third year, and Gay and Cole's (1967) results on mediation and speed of learning in illiterate Liberian children, it becomes necessary to conclude that it has not been demonstrated that Africans necessarily fall behind Western levels in late infancy or early childhood.

#### *Quality of Research*

It is not easy to evaluate the care with which the original observations were made, but it is clear that most of the studies reviewed contain flaws of design, analysis, and reporting. Certain deficiencies are evident from the above accounts and from Table 1. Sampling was unduly casual in most cases. In the studies of Langton, Ainsworth, Ramarosaona, and Lusk and Lewis, sampling even by age was adventitious, and for this reason, and for their sketchiness vis-à-vis our purposes, they may be dropped from consideration altogether. Of the rest, Faladé, Geber, and Massé combine longitudinal and cross-sectional data without separate analysis. Geber, whose studies have been in several ways the most extensive, presented findings in a fashion that is inadequate even at a descriptive level. Inferential statistics tend to be lacking. All in all, only the studies conducted by Kilbride, Theunissen, Falmagne, and Liddicoat and Koza are presented in a comprehensive, orderly, and well-analyzed manner.

#### *Adequacy of Assessment Procedures*

Half the studies used Gesell or Gesell-based scales of assessment. Knobloch (1958) criticized Geber for misinterpretation of the Gesell examination, pointing out that those without intensive experience of its administration have a constant tendency to interpret behavior at an unduly high level. Further, Cobb (1958) has suggested that the relative decline phenomenon could be an artifact of the method of Gesell DQ calculation (Mental Age/Chronological Age × 100). Clearly, quite small advances in the early months of life can give enormous DQs and constructors of other baby tests have recognized this problem and deal with it variously (Bayley, 1933, 1936, 1969; Griffiths, 1954).

The test norms themselves are unlikely to be adequate as a population, continental, or racial yardstick. Of the Gesell schedules, Cronbach (1964) said: "the standardization, reliability, and interpretation of scores are open to question." Gesell's norms were primarily based on the performance of 107 children, who were not studied intensively (Gesell & Thompson, 1938; Thomas, 1970). A proper restandardization remains to be effected. It is also easy to show that different baby tests place the average appearance of abilities and behaviors at different age levels (see Liddicoat, 1969, for instances). The Bayley test has been standardized twice with an interval of 30 years: first on a small and unrepresentative sample, and then in a most comprehensive manner. The two motor versions, essentially of norms rather than of test items, give markedly different DQs: the examples from Kilbride above show that the earlier standardization set the American average age levels higher, indicating greater Ugandan precocity on the first version.

It seems possible, therefore, that the Gesell and other poorly standardized baby tests may provide norms that contribute an artifactual element to the measurement of African infant precocity. The new Bayley test, however, has been very carefully standardized, and the most compelling evidence for African precocity based on comparison with test norms comes, therefore, from Kilbride and Goldberg, using the motor scale only, and from Lusk and Lewis (in press), who used both scales but with only 10 subjects of various ages.[8] The real problem, of course, relevant to all research except that of Theunissen and Falmagne, is that comparisons were made only with available Western norms, and not with samples of European babies taken by the same investigator. A personal element is probably inevitable in the administration of baby test items and in the use of the scoring categories, which makes it especially important that direct comparisons should be made by the same researcher. A further point that may be important is that most of those who report African precocity against norms made observations in the infants' own homes and then compared performance with normative data collected in clinic situations.

### *Two Contrary Studies*

These considerations imply that Theunissen and Falmagne must be accorded far greater weight than a couple of contradictory findings would normally merit. The only genuinely comparative investigations cannot be casually dismissed as random odd men out, especially as they are among the few well-analyzed and well-presented studies.

### *Conclusion*

The following conclusions, therefore, can be made:

1. African infant precocity has not been satisfactorily demonstrated;
2. The corresponding null hypothesis is favored by the two well-designed studies;
3. Africans have not been conclusively shown to fall systematically below Western levels in late infancy and early childhood.
4. The main source of generalizations about African infant development, the work of Geber, compares most unfavorably with at least half the other reports reviewed.

## STUDIES OF NEWBORNS

Studies at birth not only extend the time period of comparison but also come closer to a consideration of independent variables.

Geber and Dean (1957b) examined 107 (113 according to Geber & Dean, 1957a) Ugandan African singletons, delivered normally and of birth weight at least 2.5 kilograms, using the techniques developed by Thomas (Thomas, Chesni, & Dargassies, 1955). Ages at assessment ranged from 0 to

---

[8] A further study which uses the revised Bayley tests is in the pipeline. H. Leiderman (personal communication December, 1970) has recently studied the relation between perceptual and motor development and social development in a sample of Kenyan (Kikuyu) babies. Although Leiderman's data analysis is incomplete at the time of writing, he mentions evidence of Bayley DQ precocity (comparing with norms) during the first year on both the motor and the mental test.

8 days, but most were less than 48 hours. The authors concluded that these babies

> proved to be at a more advanced state of development than new-born European children. The advance was shown chiefly by a lesser degree of flexion, by a remarkable control of the head and by the frequent absence of primitive reflex activity [p. 1219].

Absence or diminution of primitive reflexes can be taken as an index of cortical control, that is, of relative maturity. Geber and Dean emphasized that much of the neonate behavior corresponded to a European age of 4–6 weeks, or even more; unfortunately, it is not stated how this judgment was formed. "The results of the examination were so consistent, and the degree of advance was so great, that there was little room for uncertainty [p. 1219]."

The lack of quantitative information in this paper is such that Geber and Dean's statements cannot be matched with their data, which in turn can scarcely be checked against other data for European newborns. Geber and Dean mentioned that they also examined 15 European and 60 Indian newborns. The results are reported in their two sentences:

> The European children gave exactly the results that have been found in Europe, and the Indian children gave almost similar results. Neither group showed any great degree of overlap with the African children [p. 1219].

It is unfortunate that a then unique study should be presented by scarcely more than verbal assurance.

Geber and Dean (1957b) placed no interpretation on their findings. A degree of vacillation is apparent in other articles. Geber and Dean (1957a) said: "it was fairly clear that the precocity had a genetic basis [p. 1061]." Subsequently, the relaxation, happiness, and continued activity of the traditional African mother-to-be are invoked in explanation (Geber, 1958a, 1958b, 1960; Geber & Dean, 1958). Geber and Dean (1964) discussed gestational age differences, nutrition, maternal attitudes and behavior, and climate as possible factors, but returned to a genetic interpretation.

Geber and Dean (1957b) reported the Moro reflex in Africans poor after 2 days and absent after 4 days (only five subjects were older than 4 days). They stated (Geber & Dean, 1957b) that the plantar response was nearly always extensor, and contrarily (Geber & Dean, 1964), that it was nearly always flexor. Griffiths (1969) has since tested both these reflexes in a cross-sectional sample of 937 African infants in Johannesburg aged between 0 and 52 weeks. Griffiths elicited the Moro in 99% of newborns and 100% in age group 1–4 weeks, extensor plantar responses in 100% at birth and up to 4 weeks. For these and for the abdominal reflex, Griffiths found development at birth and over the first year corresponding to that given for European infants by Dekaban (1959).

Parkin and Warren (1969) presented a preliminary report of a study using techniques largely derived from Prechtl and Beintema (1964) to compare 104 African newborns with 50 Europeans, aged between 36 hours and 5 days, and all born in Uganda during the same time period. The preliminary analysis, however, was based on 90 Africans and 41 Europeans. Prechtl's methods take account of the "state" that the infant is in for the assessment of each item (cf. also Bell, 1963; Escalona, 1962), an essential provision of fairly recent introduction, which is not stressed in the Thomas et al. (1955) directions or in the Geber and Dean (1957b) report. Parkin and Warren (1969) also employed more stringent sampling criteria than Geber and Dean; they accepted singletons of normal delivery and of birth weight at least 2.5 kilograms, with additional checks on the condition of the baby at birth and on gestational age by physical signs and by dates when available.

In general, it is clear that Parkin and Warren's findings differ from Geber and Dean's conclusions. These authors could see "little room for uncertainty" and neglected to calculate differences: Parkin and Warren, in contrast, found themselves unable, during the course of their research, to sustain any general impression of a difference between the groups, and were not able to note consistent differences before calculating their significance. Some examples of specific discrepancies are as follows: Geber and Dean found the doll's eye reflex completely absent in all African newborns, and the grasping and walking reflexes feeble or rare; Parkin and Warren found the doll's eye reflex present in every baby, African and European, and no group differences for these three reflexes. Geber and Dean found the Moro in African newborns poor after 2 and absent after 4 days: Parkin and Warren obtained Moro responses from every baby, the preliminary analysis indicating a higher threshold among Africans by one method of elicitation. Geber and Dean were impressed by African eye and head movements in visual pursuit of moving objects: Parkin and Warren were equally impressed by African and European subjects. However, Parkin and Warren have yet to publish the full analysis of these data, so that the detail of their conclusions is not, as yet, available.

Another preliminary report concerns Bushmen babies (possibly a distinct race). Konner (1972) summarized the examination of eight newborns according to the Prechtl-Beintema scheme: In form and degree the reflex repertoire of Bushmen babies is the same as their European counterparts. Konner noted none of the specific departures from the European pattern reported by Geber and Dean. This study also awaits publication in full.

#### *Studies at One Month*

In connection with the neonatal period, it is worth mentioning findings reported for one month of age. Geber (e.g., 1957a) spoke of a remarkable precocity from the first month, but gave no scores. Faladé (1955) found an average Gesell DQ of 163 for 13 infants. Kilbride tested but three children: the Bayley revised DQ averaged 120; the 1936 Bayley DQ averaged 174. Theunissen (see Footnote 2) and Falmagne (1962) again provided the genuine comparisons and the exceptions. At one month, Falmagne's European sample was superior in 7 items, his African sample on 2 items, 11 items affording no distinctions. Theunissen's 4-week Gesell schedule shows that Europeans were superior on 6 items, Africans on 4 items, with 1 item equal.

### *Is Precocity Present at Birth?*

#### *Findings*

No tally is feasible. Geber and Dean (1957b), for so long the only study available, have suggested to Brackbill and Thompson (1967), Zigler and Child (1969), and Levine (1970) that African precocity has a hereditary basis. Geber and Dean themselves derived this conclusion in some publications but not others. Weyl and Possony (1963), Jensen (1969), and Eysenck (1971) further inferred from hereditary early precocity the likelihood of genetically determined African intellectual inferiority at full maturity. All of these authors accept Geber and Dean's findings uncritically.

#### *Quality of Research*

Comments are largely confined to Geber and Dean (1957b), most of whose deficiencies are indicated in the above account. Of the subjects 78% were less than 48 hours old, an age advised against by Prechtl and Beintema (1964). Consideration of the "state" of the baby, necessary for the evaluation of each item, is nowhere mentioned. The most serious criticism, however, is that the single quantitative affirmation, that Ugandan neonatal behavior mostly corresponded to a European age of 4–6 weeks, or even more, is mysteriously derived. We know of no body of developmental information that would permit such a comparison, for most neonatal assessment items. And the comparison cannot have been with the 15 European subjects examined, since these were also newborns; in any case, the European babies are mentioned at the end of the article, well after the conclusion of remarkable African precocity. Detailed comparison of these 15 with the African sample would have been a valuable addition to the study.

The major methodological point parallels that of the previous discussion section: Comparison of newborn groups must be made by the same researchers operating with the same methods over the same period of time. Were extensive norms available for neonatal behavior, it would remain that the examination is a difficult business, that control or modification of "state" is a matter of skill, that item order is dependent on inspiration, and that some scoring categories are fuzzy at the edges. Only Parkin and Warren's design takes account of this requirement.

### *Relevance of Neonate Measures*

One further point should be mentioned, if only in passing. We have no good indication that the neonate measures used are on a developmental continuum with those capacities assessed by a baby test or other measures used at later ages. Scales of assessment such as those of Thomas or Prechtl have the primary medical purpose of diagnosing pathology. Insofar as they validly measure central nervous system pathology, they may afford some prognosis of later development. But as measures of differences within the normal range, they may be quite irrelevant to subsequent development. There seems to be no good predictive evidence that this is not the case (Thomas, 1970).

### *Conclusion*

Judgment must be reserved concerning African precocity at birth. The major published report to date is unconvincing in itself, and, though three other studies suggest further reservations, they are as yet too limited or too preliminary to allow us any greater conclusiveness.

## DIFFERENCES BY SOCIAL LEVEL

Differences by social level or milieu have not been emphasized in the preceding accounts. Vouilloux (1959) found orphanage babies retarded, averaging 78 DQ. This is within the range expected by simply inductive generalization from Western studies of institutionalized babies. Differences by social level among African babies are, however, not similarly predictable.

Geber (1958b) tested 60 infants from "Westernized" homes of a superior social level. Ten or more children in each of the first 5 years of age were tested cross-sectionally; the Gesell test was supplemented by the Terman-Merrill test from age 2. Development in the first 2 years was summarized as somewhat precocious, but not as remarkably so as in the poorer families. However, the relative decline did not occur to nearly the same extent. The only figures given (Geber, 1960) show that 59 of the 60 children obtained Gesell DQs of 100 or above and that 32 of the 40 children tested with the Terman-Merrill had IQs of 100 or greater. Average quotients were neither given nor calculable, but Geber and Dean's (1964, p. 433) graph provided a separate curve for "Westernized" children, showing a DQ which over 6 years appears to stay within the 110–120 range, and thus achieves a superiority to that of the poorer children from the middle of the second year.

Ramarosaona (see Footnote 5) also claimed that the few infants of a higher social level in her group paralleled the development of Geber's elite babies. And Vouilloux (1959) was able to divide her sample of 48 healthy infants into 21 of a low socioeconomic level and 27 of a medium or high level. She found a highly significant difference between the mean DQs, which are given as 130.9 for the former and 114.9 for the latter group.

### *Are There Differences by Social Level?*

### *Tally*

For once, the results, sparse as they are, show consistency. Infants from homes of a low social level are initially more advanced than elite babies. Only Geber reported on the differential decline beyond infancy.[9]

### *Quality of research*

Some of the earlier criticisms of these researchers apply again here, and Ramarosaona's (see Footnote 5) work can be considered too sketchy to carry any real weight. However, the major criticism leveled, that comparison with norms alone is necessarily inadequate, does not apply, since two African groups were compared by the same researcher. A bias, either in the Gesell norms or in the testing procedures, cannot account for the differences reported by social level. Moreover, the difference is the actual reverse of what might be expected from Western assumptions and from knowledge of African environments of babies.

---

[9] A study of Nigerian (Yoruba) children by Abiola (1965) should be mentioned, but is difficult to fit in because of the poor design and self-contradictory report. Abiola tested elite and traditional-urban (Oje) children cross-sectionally from 1 to 5 years, using a perceptual-motor test and a conceptual-verbal test, both of his own construction but drawing items from Western tests and scales. As the material was not graded by age, the 1-year-old children were incapable on most items, and the possibility of ceiling effects at the higher ages is a strong one. Abiola stated that the traditional children were superior on the perceptual-motor test at 1 year of age, but his table of results indicates the opposite. Significant differences favoring the elite children were reported for 3 and 5 years on the perceptual-motor test, and for 2 and 5 years on the conceptual-verbal test. Abiola's suggestion that a convergence of elite and traditional performance takes place between 3 and 5 years has been investigated by Lloyd (1971) in an excellent study. Lloyd used a modified Stanford-Binet, concrete operational tasks, and a relational concept-learning problem to assess Yoruba elite and traditional children cross-sectionally, from 3½ to 8 years, and Abiola's convergence hypothesis was not supported.

### *Conclusion*

It can hardly be claimed that we have here a definite phenomenon. However, there is a somewhat surprising indication that infants from poor African homes develop more rapidly at first than those from elite homes.

## DISCUSSION

One should now ask, Is this research issue worth taking further? And if so, how?

### *African Precocity*

African precocity was considered undemonstrated. The two truly comparative studies dominate the picture beyond the newborn period, and judgment must be reserved on differences at birth. Any further comparisons must include actual samples of non-African infants. They should also introduce more suitable methods of assessment: the scales for newborns and the standard baby tests all tend toward a quick and unreliable sample of conveniently observed behavior, and thus present a largely atheoretical hotchpotch of rather superficial items. Newborn comparisons would profit from more "psychological" measures, for example, behavioral scales as in Brazelton, Robey, and Collier (1969), and careful observations of perception, learning, and reactivity. The longitudinal relation between neonatal status and subsequent development would show whether variation in newborns other than that diagnostic of pathology is merely irrelevant.

The standard baby tests are useful for first approximations, but otherwise must be replaced by more sophisticated, careful, and intensive approaches. The introduction to cross-cultural research of clinical-experimental longitudinal study of the Piagetian kind, or of the ordinal scales of infant assessment based on Piaget's (1953) theory of sensorimotor development, would be timely. Other recent developments, for example, White (1967), Bruner (1968), Schaffer and Parry (1969), and linguistic approaches, are also relevant; it is to be hoped that Konner's (1972) ethological studies of Bushman babies will suggest novel lines of investigation.

### *Research Questions*

To focus directly on "African precocity," simply to compare African and non-African baby samples, however well taken, may be unnecessarily naive. The clear prima facie reason for expecting uniform African development is a racial-genetic hypothesis. Weyl and Possony (1963), Jensen (1969), and Eysenck (1971) uncritically accepted Geber's reports of precocity and argued from them to the likelihood of hereditary racial differences in mature intelligence. In doing so, they appear to assume knowledge of the causal linkages between precocity at one level and retardation at another level of development—knowledge that we do not yet have. Their real problem, however, in the light of this review, is that the superstructure of their argument rests on the phenomenon of African infant precocity as evidenced by Geber. It cannot rest on the conclusions drawn here. Racial-genetic hypothesis, as Jensen insisted, cannot be excluded from consideration, but further research is necessary if it is to be supported by African data. Such research might well include straightforward comparisons between African and non-African samples, particularly for the neonatal period; but ultimately the racial-genetic question would have to be unpacked into more sophisticated research strategies, for example, by studying African foster-children in European homes.

No other hypothesis would plausibly suggest uniform African development across the continent. Uniformity of infant environments is too much to ask. Cravioto (Cravioto, 1968; Cravioto, Birch, De Licardie, Rosales, & Vega, 1969; Cravioto, De Licardie, & Birch, 1966), citing Gesell-based studies from Mexico and Guatemala in conjunction with the work of Geber, proposed that the precocity of infants in "preindustrial" cultures may be a general phenomenon, not limited to Africa (cf., however, Brazelton et al., 1969). If Cravioto, though vague, is right, precocity would be found in many samples of African babies. The proposal has testable aspects: "preindustrial" cultures are to be found in many parts of the world, regardless of racial boundaries; but also "modern" or "acculturated" groups are to be found in Africa. Thus the finding we have noted of African differences by social level assumes importance as an empirical lead.

A high incidence of disease and undernutrition, multiple caretaking, a lack of obvious toys or playthings, an unfavorable signal-to-noise ratio—these and other "deficiencies" are likely to be the African infant's lot in a poor home. In Western countries, social class differences in baby-test performance up to 18 months or more have been generally nonexistent or negligible (Bayley, 1965; Golden & Birns, 1969; Hindley, 1960; Knobloch & Pasamanick, 1960). Kagan (1966) found a lower-class deficit in cognitive functioning by 12 months of age. This contrast between Africa and the West may provide a focal point for developing alternatives to current theory and assumption. Milieu comparisons within African societies could include orphanages and urban-slum homes as well as traditional-rural and elite backgrounds.

The purpose of investigating milieu differences would be to break down the global and vague conceptions of milieux by empirical measurement into those aspects of care, rearing, and experience that relate to infant development, thus eliminating milieu concepts as research variables in favor of more precise influences. This would seem the most sensible way of closing on independent variables, which have tended to be left to speculation so far, and would probably be a largely inductive, dragnet procedure. Hypotheses would not be lacking: Rosenzweig (1966), for example, favored early stimulation as the main cause, discussing Geber's work in connection with animal studies showing that learning ability increases with greater variety and richness of experience; Walters and Parke (1965) pointed to Geber's findings as an instance of the role played by the distance receptors in development, implicating the environmental variability provided for the African infant. However, granted that an inductive approach still necessitates selectivity, this variety of the cross-cultural strategy has the primary goal of discovering something, of producing a theory. Thus retention of a major dragnet element in pinpointing independent variables seems necessary. If precise independent variables of infant stimulation and experience prove identifiable, the research category of "African" may be appropriately discarded, exit may be made from the cross-cultural cycle, and further testing may proceed in an intracultural, even experimental, fashion.

## SUMMARY AND CONCLUSIONS

A review of research on African infant development led to these views: (*a*) African infant precocity cannot be taken as an acceptably established empirical phenomenon; (*b*) African differences reported by social milieu afford a more sensible basis for further enquiry.

Discussion generated the following proposals: (*a*) More sophisticated methods of infant assessment and study should be employed in this kind of research; (*b*) comparative status at birth needs further investigation, especially in terms of psychologically relevant variables and with follow-up beyond the neonatal period; (*c*) improved research designs are necessary to identify independent variables that account for differences in development. The cross-cultural cycle for this research issue is as yet incomplete. Further study should aim to ensure that this variant of the cross-cultural approach will serve its true purpose by rendering itself obsolete.

## REFERENCES

ABIOLA, E. T. The nature of intellectual development in Nigerian children. *Teacher Education,* 1965, 6, 37–58.

AINSWORTH, M. D. S. The development of infant-mother interaction among the Ganda. In B. M. Foss (Ed.), *Determinants of infant behaviour.* Vol. 2. London: Methuen, 1963.

AINSWORTH, M. D. S. *Infancy in Uganda.* Baltimore, Md.: John Hopkins University Press, 1967.

BARDET, C., MASSÉ, G., MOREIGNE, F., & SENECAL, J. Application du test de Brunet-Lézine à un groupe d'enfants Ouolofs de 6 mois à 24 mois. *Bulletin de la Société Médicale d'Afrique Noire de la Langue Française,* 1960, **5**, 334–356.

Bayley, N. *The California first-year mental scale.* Berkeley, Calif.: University of California Press, 1933.

Bayley, N. *The California infant scale of motor development.* Berkeley, Calif.: University of California Press, 1936.

Bayley, N. Comparisons of mental and motor test scores for ages 1–15 months by sex, birth order, race, geographical location and education of parents. *Child Development,* 1965, **36**, 379–411.

Bayley, N. *Bayley scales of infant development.* New York: Psychological Corporation, 1969.

Beadle, M. *A child's mind.* New York: Doubleday, 1970.

Bell, R. Q. Some factors to be controlled in studies of the behaviour of newborns. *Biologia Neonatorum,* 1963, **5**, 200–214.

Brackbill, Y., & Thompson, G. G. (Eds.) *Behavior in infancy and early childhood.* New York: Free Press, 1967.

Brazelton, T. B., Robey, J. S., & Collier, G. A. Infant development in the Zincanteco Indians of Southern Mexico. *Pediatrics,* 1969, **44**, 274–293.

Bruner, J. S. *Processes of cognitive growth: Infancy.* Worcester, Mass.: Clark University Press, 1968.

Casler, L. Maternal deprivation: A critical review of the literature. *Monograph of the Society for Research in Child Development,* 1961, **26** (Serial No. 2).

Cobb, J. C. Precocity of African children. *Pediatrics,* 1958, **22**, 601–604.

Cravioto, J. Nutritional deficiencies and mental performance in childhood. In D. C. Glass (Ed.), *Environmental influences.* New York: Rockefeller University Press, 1968.

Cravioto, J., De Licardie, E. R., & Birch, H. G. Nutrition, growth, and neurointegrative development: An experimental and ecologic study. *Pediatrics,* 1966, **38**, 319–372.

Cravioto, J., Birch, H. G., De Licardie, E., Rosales, L., & Vega, L. The ecology of growth and development in a Mexican preindustrial community. *Monograph of the Society for Research in Child Development,* 1969, **34** (5, Serial No. 129).

Cronbach, L. J. *Essentials of psychological testing.* New York: Harper, 1964.

Dekaban, A. *Neurology of infancy.* Baltimore: Williams & Wilkins, 1959.

Ebert, E., & Simmons, K. The Brush Foundation study of child growth and development: 1. Psychometric tests. *Monograph of the Society for Research in Child Development,* 1943 **8** (2, Serial No. 35).

Escalona, S. K. The study of individual differences and the problem of state. *Journal of the American Academy of Child Psychiatry,* 1962, **1**, 11–37.

Eysenck, H. J. *Race, intelligence and education.* London: Temple Smith, 1971.

Faladé, S. *Contribution à une étude sur le développement de l'enfant d'Afrique noire.* Paris: Foulon, 1955.

Faladé, S. Le développement psycho-moteur de l'enfant africain du Sénégal. *Concours Médical,* 1960, **82**, 1005–1013.

Falmagne, J-Ce. Étude comparative du développement psychomoteur. *Académie Royale des Sciences d'Outre-Mer, Mémoires,* 1962, **13** (5).

Gay, J., & Cole, M. *The new mathematics and an old culture.* New York: Holt, 1967.

Geber, M. Développement psycho-moteur de l'enfant africain. *Courrier,* 1956, **6**, 17–29.

Geber, M. The psychomotor development of African children in the first year, and the influence of maternal behaviour. *Journal of Social Psychology,* 1958, **47**, 185–195. (a)

Geber, M. L'enfant africain occidentalisé et de niveau social supérieur en Ouganda. *Courrier,* 1958, **8**, 517–523. (b)

Geber, M. Problèmes posés par le développement du jeune enfant africain en fonction de son milieu social. *Travail Humain,* 1960, **1**, **2**, 97–111.

Geber, M. Longitudinal study and psychomotor development among Baganda children. In, *Proceedings of the Fourteenth International Congress of Applied Psychology.* Vol. 3. Copenhagen: Munksgaard, 1962.

Geber, M., & Dean, R. F. A. Gesell tests on African children. *Pediatrics,* 1957, **20**, 1055–1065. (a)

Geber, M., & Dean, R. F. A. The state of development of new born African children. *Lancet,* 1957, **1**, 1216–1219. (b)

Geber, M., & Dean, R. F. A. Psychomotor development in African children: The effects of social class and the need for improved tests. *Bulletin of the World Health Organisation,* 1958, **18**, 471–476.

Geber, M., & Dean, R. F. A. Le développement psycho-moteur et somatique des jeunes enfants africains en Ouganda. *Courrier,* 1964, **15**, 425–437.

Gesell, A., & Thompson, H. *The psychology of early growth.* New York: Macmillan, 1938.

Golden, M., & Birns, B. Social class and cognitive development in infancy. *Merrill-Palmer Quarterly,* 1969, **14**, 139–149.

Griffiths, J. McE. Development of reflexes in Bantu children. *Developmental Medicine and Child Neurology,* 1969, **11**, 533–535.

Griffiths, R. *The abilities of babies.* London: University of London Press, 1954.

Hindley, C. B. The Griffiths scale of infant development: Scores and predictions from 3 to 18 months. *Journal of Child Psychology and Psychiatry,* 1960, **1**, 99–112.

Hudson, W. The study of the problem of pictorial perception among unacculturated groups. *International Journal of Psychology,* 1967, **2**, 90–107.

Jahoda, G. Problems in African education. *Journal of Social Issues,* 1968, **24**, 161–178.

Jensen, A. R. How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review,* 1969, **39**, 1–123.

Kagan, J. A developmental approach to conceptual growth. In H. J. Klausmeier & C. W. Harris (Eds.), *Analyses of concept learning*. New York: Academic Press, 1966.

Kagan, J. Personality development. In I. L. Janis, G. F. Mahl, J. Kagan, R. R. Holt (Eds.), *Personality: Dynamics, development and assessment*. New York: Harcourt, 1969.

Kessell, F. S. The philosophy of science as proclaimed and science as practiced: "Identity" or "dualism"? *American Psychologist*, 1969, **24**, 999–1005.

Kilbride, J. E., Robbins, M. C., & Kilbride, P. L. The comparative motor development of Baganda, American white, and American black infants. *American Anthropologist*, 1970, **72**, 1422–1427.

Knobloch, H. Precocity of African children. *Pediatrics*, 1958, 22, 601–604.

Knobloch, H., & Pasamanick, B. Environmental factors affecting human development, before and after birth. *Pediatrics*, 1960, **26**, 210–218.

Konner, M. J. Aspects of the developmental ethology of a foraging people. In N. G. Blurton-Jones (Ed.), *Ethological studies of child behaviour*. London: Cambridge University Press, 1972.

Langton, E. A. C. Some observations of infants and young persons in Bunyoro, Uganda. *East African Medical Journal*, 1934–1935, **11**, 316–323.

LeVine, R. A. Cross-cultural study in child psychology. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology*. New York: Wiley, 1970.

Liddicoat, R. Development of Bantu children. *Developmental Medicine and Child Neurology*, 1969, 11, 821.

Liddicoat, R., & Koza, C. Language development in African infants. *Psychologia Africana*, 1963, **10**, 108–116.

Lloyd, B. B. The intellectual development of Yoruba children: A re-examination. *Journal of cross-cultural psychology*, 1971, **2**, 29–38.

Lusk, D., & Lewis, M. Mother-infant interaction and infant development among the Wolof of Senegal. *Human Development*, in press.

Massé, G. *Croissance et développement de l'enfant en Dakar*. Paris: Centre Internationale de l'Enfance, 1969.

Moreigne, F., & Senecal, J. Résultats d'un groupe d'enfants africains au Terman-Merrill. *Revue de Psychologie appliquée*, 1962, **12**, 15–32.

Parkin, J. M., & Warren, N. A comparative study of neonatal behaviour and development. *Proceedings of University Social Science Conference*. Nairobi, Kenya: University of East Africa, 1969.

Piaget, J. *The origins of intelligence in the child*. London: Routledge, 1953.

Prechtl, H., & Beintema, D. *The neurological examination of the full term newborn infant*. London: Heinemann, 1964.

Rosenzweig, M. R. Environmental complexity, cerebral change, and behaviour. *American Psychologist*, 1966, **21**, 321–332.

Schaffer, H. R., & Parry, M. H. Perceptual-motor behaviour in infancy as a function of age and stimulus familiarity. *British Journal of Psychology*, 1969, **60**, 1–9.

Sontag, L. W., Baker, C. J., & Nelson, V. L. Mental growth and personality development: A longitudinal study. *Monograph of the Society for Research in Child Development*, 1958, 23 (68, Serial no. 2).

Terman, L. M., & Merrill, M. A. *The Stanford-Binet Intelligence Scale*. New York: Houghton Mifflin, 1960.

Thomas, A., Chesni, Y., & Dargassies, S.-A. *Examen neurologique du nourrisson*. Paris: La Vie Médicale, 1955.

Thomas, H. Psychological assessment instruments for use with human infants. *Merrill-Palmer Quarterly*, 1970, **16**, 179–224.

Thompson, W. R. Development and the biophysical bases of personality. In E. F. Borgatta & W. W. Lambert (Eds.), *Handbook of personality theory and research*. Chicago: Rand McNally, 1968.

Vouilloux, Dr. Étude de la psycho-motricité d'enfants africains au Cameroun: Test de Gesell et réflexes archaiques. *Journal de la Société des Africanistes*, 1959, **29**, 11–18.

Walters, R. H., & Parke, R. D. The role of distance receptors in the development of social responsiveness. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior*. Vol. 2. New York: Academic Press, 1965.

Weyl, N., & Possony, S. *The geography of intellect*. Chicago: Regnery, 1963.

White, B. An experimental approach to the effects of experience on early human behavior. In J. P. Hill (Ed.), *Minnesota Symposia on Child Psychology*. Vol. 1. Minneapolis: University of Minnesota Press, 1967.

Zigler, E., & Child, I. L. Socialization. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology*. Reading, Mass.: Addison Wesley, 1969.

(Received April 7, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 5, 368-374

# TYPE IV ERRORS AND INTERACTIONS

JOEL R. LEVIN[1]
*University of Wisconsin*

LEONARD A. MARASCUILO
*University of California, Berkeley*

Marascuilo and Levin's (1970) notion of Type IV errors is extended, with particular attention paid to the interpretation of interactions in factorial analysis of variance (ANOVA) designs. To help clarify what an interaction is and what it is not in terms of the ANOVA model, it is emphasized that (*a*) comparisons that are designed to explain an interaction are acceptable only if they reduce to comparisons involving interaction parameters exclusively; and (*b*) such comparisons may be both specified and directional, in that they may be defined to test an investigator's a priori hypotheses.

The rejection or denial of a true statistical hypothesis and the nonrejection or acceptance of a false hypothesis are abstractions with which researchers in the behavioral sciences generally make only passing acquaintance. While one is taught to express concern for—and even to compute probabilities associated with—Type I and Type II errors as a student attending a first course in statistics, the same student in the "cruel world" of research experiences understandable difficulty in deciding when either of these errors has in fact occurred. As a result, one gradually learns to live with them. Such should not be the case with Type IV errors, as introduced by Marascuilo and Levin (1970), since with practice this kind of error is easily recognized and consequently avoided.[2]

According to the definition of Marascuilo and Levin, a Type IV error is said to occur whenever a correct statistical test has been performed, but is then followed by analyses and explanations that are not related to the statistical test used to decide whether the hypothesis should or should not have been rejected. More succinctly, a Type IV error is made whenever a researcher offers an incorrect interpretation to a correctly rejected statistical hypothesis. Less succinctly, a Type IV error is identified as having been committed whenever a researcher concludes, on the basis of an appropriately performed statistical test, that there is a reliable source of variability in the data, but then proceeds to specify the locus of the effect with an eyeball interpretation of the data or by employing post hoc multiple comparison procedures that are not congruent with the hypothesis initially tested and that may not even correspond to the underlying model upon which the statistical test was based.[3]

## *Type IV Errors in the One-Way Analysis of Variance Model*

Perhaps the most commonly encountered Type IV error is the one committed by a researcher who follows a rejected analysis of variance (ANOVA) hypothesis with a set of overlapping multiple $t$ tests, each performed at the same alpha level as chosen for the original $F$ test. In this case, the statistical hypothesis $H_0\colon \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_I$ has been rejected with the probability of a Type I error set equal to $\alpha$. If the researcher now examines each of the $Q = \frac{I(I-1)}{2}$ paired mean comparisons at the same $\alpha$ level as used for the $F$ test, the total probability of at least one Type I error in the *set* of comparisons is inflated far above the original probability to a maximum value of

---

[1] This article represents a substantially reduced version of a paper which was presented at the annual meeting of the American Educational Research Association in New York, March 1971.

Requests for reprints and for copies of the original paper should be sent to Joel R. Levin, Wisconsin Research and Development Center, 1404 Regent Street, Madison, Wisconsin 53706.

The authors are grateful to Lawrence Hubert, Maryellen McSweeney, Neil H. Timm, G. William Walster, and Charles E. Woodson, whose helpful comments have been incorporated.

[2] Readers who are still wondering about Type III errors may refer to the assorted definitions reviewed by Marascuilo and Levin (1970).

[3] In this discussion, the terms "comparison" and "contrast" are used interchangeably to refer to a linear combination of cell means for which the sum of the selected coefficients equals zero.

$\alpha_0 \leq \alpha Q$. Should this procedure pronounce statistically significant certain pairwise comparisons that would not have been identified with the "appropriate" post hoc Scheffé (1953) method, then a Type IV error would have been made.[4]

Some researchers may attempt to control the maximum probability of at least one Type I error in the set of $Q$ pairwise contrasts (or in the set of any $K$ planned comparisons, i.e., contrasts involving linear combinations of means, as well as pairwise comparisons) by using Bonferroni critical values as described by Miller (1966) or Dunn (1961) to conduct their post hoc investigations. However, even these adjustments do not eliminate Type IV errors since they are not related to the original $F$ test in a one-to-one manner. Because the Bonferroni procedures are quite powerful—especially when the number of comparisons of interest is relatively small—they may detect a greater number of significant differences than would the Scheffé method, which corresponds *exactly* to the classical $F$ test.

It is worth mentioning that Type IV errors of this kind may be avoided simply by bypassing the $F$ test altogether. In point of fact, abandoning the $F$ test might be the optimal strategy for a researcher to follow if the plan is to examine only a small number of contrasts, since the $F$ test could lead to a nonrejection of $H_0$ while one or more comparisons could be identified as significant with the more powerful Bonferroni or Dunn method. In such cases, the relative efficiency of the Scheffé to the Dunn procedure (defined, perhaps, in terms of the ratio of the respective critical values) may be determined prior to data collection, in order to reach a rational decision concerning the approach to adopt (see Davis, 1969, or Kirk, 1968).[5]

It is also worth noting that in the equal sample size model, employing the Tukey method of pairwise comparisons following the rejection of $H_0$ based on the classical $F$ test may also produce Type IV errors, since Tukey's procedure is more likely to identify pairwise differences as significant than is Scheffé's (see Scheffé, 1959). The reason for this potential discrepancy is that the two procedures are derived from different underlying distributions. Tukey's more powerful method is based on the distribution of the studentized range for which the corresponding test statistic is the ratio of the weighted (by the square root of the common sample size) maximum mean difference to the square root of the mean square within. Scheffé's procedure, on the other hand, is based on a different mathematical model for which the test statistic is the familiar ratio of the mean square between to the mean square within in the one-way ANOVA model. Thus, if a researcher is interested in performing only pairwise contrasts, and if sample sizes are equal, then the $F$ test is not the "appropriate" test to perform! In this case, the researcher should first perform the studentized range test and if it leads to rejection of $H_0$, then Tukey's method of pairwise comparisons should be employed (Dixon & Massey, 1969). With this strategy, the probability of making a Type IV error is reduced to zero.

## *Type IV Errors in Other Models*

Cautions regarding the incorrect post hoc analysis of correctly performed omnibus tests are not confined to traditional ANOVA designs. Marascuilo (1966) has described simultaneous inference procedures that are "appropriate" for large-sample tests of the differences among $J$ independent proportions and among $J$ independent correlation coefficients. Steel (1961), Dunn (1964), and Marascuilo and McSweeney (1967) have developed multiple comparison techniques to accompany the nonparametric rank tests. The essential feature of such post hoc methods is that they are *based on the same distribution as the test statistic*, and therefore will yield information that is congruent with the test initially performed.

As might be surmised, multivariate ANOVA hypotheses offer the researcher a Pandora's box of very elegant and sophisticated Type IV errors (among others) that may result in unwarranted explanations of significant findings. One common error following a rejected multivariate hypothesis is to perform variable-by-variable comparisons (perhaps at a reduced $\alpha$ level, or using post hoc univariate techniques), or to interpret the significant multivariate statistic in terms of linear combinations of dependent variables, as might be suggested from an examination of principal components or linear discriminant functions. While there may be some correspondence between the decisions made under these analysis procedures and the "appropriate" Roy-Bose multivariate post hoc method as described by Morrison (1967), it has been shown by Hummel and Sligo (1971) that the Type I error probabilities of the multivariate and univariate procedures are not identical. That is, when multivariate data are analyzed on a post hoc basis with critical values determined from univariate procedures, one is liable to arrive at statistical decisions that are different from those based on multivariate post hoc techniques. As was mentioned for the univariate case, if only a small number of comparisons are of interest, it might be advisable to examine those comparisons individually rather than to perform an overall multivariate $F$ test. In support of this notion, it has recently been shown that the Bonferroni method produces uniformly narrower intervals under the null distribution than the usual multivariate post hoc procedures, even for large numbers of planned contrasts.[6]

---

[4] "Appropriate" is used advisedly here, in the sense that the Scheffé procedure is the *only* procedure that corresponds *exactly* to the initial test of hypothesis. Whether Scheffé's procedure is desirable (e.g., with respect to statistical power) is another issue which has been discussed elsewhere (e.g., Petrinovich & Hardyck, 1969).

[5] This procedure should be interpreted with caution, however, since it involves comparing critical values of null distributions based on the same sample size rather than comparing the respective *powers* of different techniques (i.e., the probabilities of rejecting the hypothesis given various alternative distributions).

Until now, the discussion of Type IV errors has been basically from a Type I error point of view. In other words, situations were described in which the correspondence between the Type I error probability of the post hoc analyses and the test initially performed was less than perfect. However, another (perhaps more serious) Type IV error occurs when a researcher defines his post hoc contrasts in a manner that does not even investigate the hypothesis he is presumably testing. This kind of error frequently occurs when it comes to interpreting statistical "interactions" in contingency tables (Goodman, 1964; Marascuilo, 1966), regression analyses,[7] and factorial ANOVA designs (Marascuilo & Levin, 1970).

Marascuilo and Levin pointed out that a typical strategy following the detection of a significant interaction in a factorial ANOVA is to make either pairwise or nested comparisons using the various cell means of the design. They showed by examples that this procedure does not correspond to the interaction $F$ test initially performed. It was further suggested that this error arises because many researchers do not have a clear understanding of what constitutes an interaction as it is defined by the mathematical ANOVA model. As a result, the post hoc procedures and/or verbal discussion based on the identification of a significant interaction often are inappropriate. A further clarification of the meaning of a statistical interaction is presented in terms of the factorial ANOVA model.

### *Mathematical Model for an I by J Factorial Design*

Consider a two-way fixed effects ANOVA model with an equal number of observations per cell.[8] For $I$ rows and $J$ columns, the model may be written as:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

where

$y_{ijk}$ = the value of the $k$th observation in the $i$th row and $j$th column;

$\mu$ = the grand mean;

$\alpha_i$ = the effect of Level $i$ for Factor A where $\sum_{i=1}^{I} \alpha_i = 0$;

$\beta_j$ = the effect of Level $j$ for Factor B where $\sum_{j=1}^{J} \beta_j = 0$;

$\gamma_{ij}$ = the joint effect of Level $i$ and Level $j$ where $\sum_{i=1}^{I} \gamma_{ij} = \sum_{j=1}^{J} \gamma_{ij} = 0$;

---

[6] See G. J. Schluck's paper, "A comparison of three confidence interval procedures for the general linear model," presented at the annual meeting of the American Educational Research Association, New York, March 1971.

[7] N. H. Timm's forthcoming book, *Multivariate Analysis with Applications in Education and Psychology*, includes a discussion of this.

[8] Although the discussion is based on the assumption of equal cell $n$'s, the same general principles may be extended to designs with unequal cell frequencies.

$\epsilon_{ijk}$ = the error associated with the $k$th observation in the $ij$th cell.

The $\epsilon_{ijk}$ are assumed to be statistically independent, normally distributed, with a mean of zero and a variance equal to $\sigma^2$.

Under this model it is customary to test for the presence of row effects, column effects, and interaction effects by means of three orthogonal tests of hypothesis:

$$H_{01}\colon \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

$$H_{02}\colon \beta_1 = \beta_2 = \cdots = \beta_J = 0$$

$$H_{03}\colon \gamma_{11} = \gamma_{12} = \cdots = \gamma_{IJ} = 0.$$

If any of the main effects is significant, one may use Scheffé's method to identify possible sources of variance in exactly the same manner as that used for the one-way ANOVA. However, if the interaction test leads to a rejection of $H_{03}$, slightly different procedures must be used to locate the significant sources that account for the rejection of the hypothesis.

As suggested by Marascuilo and Levin (1970), the common practice of making simple comparisons among the cell means to locate the sources of a significant interaction is not valid if $H_{03}$ is rejected. The procedure is appropriate only if instead of testing $H_{01}$, $H_{02}$, and $H_{03}$ as orthogonal hypotheses, one were to test the composite total cell hypothesis:

$$H_{04}\colon \mu_{11} = \mu_{12} = \cdots = \mu_{IJ}.$$

In this case, the Scheffé coefficient would be given by:

$$S = \sqrt{(IJ-1)F_{(IJ-1),IJ(n-1)}(1-\alpha)}$$

where $n$ = number of observations per cell, and $\alpha$ = the probability of a Type I error associated with $H_{04}$.[9] In this case, a typical pairwise contrast is defined by:

$$\begin{aligned}\Psi &= \mu_{ij} - \mu_{i'j'}\\ &= (\mu + \alpha_i + \beta_j + \gamma_{ij})\\ &\qquad - (\mu + \alpha_{i'} + \beta_{j'} + \gamma_{i'j'})\\ &= (\alpha_i - \alpha_{i'}) + (\beta_j - \beta_{j'}) + (\gamma_{ij} - \gamma_{i'j'}).\end{aligned}$$

Note, however, that if this contrast were found to be significant, it would be impossible to know whether or not the difference was due to the fact that $\alpha_i \neq \alpha_{i'}$, or that $\beta_j \neq \beta_{j'}$, or that $\gamma_{ij} \neq \gamma_{i'j'}$, or any combination of these. In words, contrasts of this type lead to a confounding of the model's parameters, and the situation is not alleviated simply by testing $H_{04}$ instead of $H_{03}$.

On the basis of this discussion it should not be concluded that there is no appropriate way to interpret the meaning of a significant $F$ ratio for interaction, in terms of some linear combination of cell means. What the previous discussion is meant to suggest is that linear contrasts of the form $\Psi = \mu_{ij} - \mu_{i'j'}$ that are typically defined by researchers to interpret interactions are incorrect in a Type IV error sense.

If $H_{03}$, the overall test of interaction, is performed and if the hypothesis is rejected, then Scheffé's method will guarantee simultaneous Type I error protection for all contrasts investigated, but only if (*a*) valid interaction contrasts are defined, along with their appropriate standard errors; and (*b*) the Scheffé coefficient is based on the degrees of freedom associated with the test of $H_{03}$. One valid interaction contrast is of the form $\Psi = \Delta_i - \Delta_{i'} = (\mu_{ij} - \mu_{ij'}) - (\mu_{i'j} - \mu_{i'j'})$, which compares the differences between two same row means in different columns. Contrasts of this kind must include a *minimum* of four cell means in order for $\gamma_{ij}$ parameters exclusively to be estimated (see Marascuilo & Levin, 1970). Other valid post hoc procedures for interpreting interactions include tests of interaction effects being equal to zero ($\gamma_{ij} = 0$) or to one another ($\gamma_{ij} = \gamma_{i'j'}$), as well as any contrast among cell means which involves the $\gamma_{ij}$ only. For each of the procedures just mentioned, the Scheffé coefficient is given by:

$$\mathrm{S} = \sqrt{(I-1)(J-1)F_{(I-1)(J-1),IJ(n-1)}(1-\alpha)}.$$

Formulas for determining the proper standard

---

[9] It is assumed that in employing a factorial ANOVA model and testing each source of variance separately, the investigator is interested in controlling $\alpha$ for the *family* of comparisons associated with each source (Miller, 1966). That is, he has formulated a priori hypotheses to accompany each of the main effects and interactions that he tests and, therefore, each corresponds more or less to a separate experiment (as is generally desired of factorial designs and which is defensible when the denominator's degrees of freedom are large). If one is concerned that the probability of a Type I error for the single *experiment* consisting of $IJ$ cells not exceed $\alpha$, then one should either test the total between-cell hypothesis with $\alpha$ or distribute $\alpha$ among those sources of variance of interest.

TABLE 1

Two-Factor Design of Sex by Age

| Age (years) | Sex | |
|---|---|---|
| | Girls | Boys |
| 6 | $\bar{y}_{11}$ | $\bar{y}_{21}$ |
| 8 | $\bar{y}_{12}$ | $\bar{y}_{22}$ |
| 10 | $\bar{y}_{13}$ | $\bar{y}_{23}$ |
| 12 | $\bar{y}_{14}$ | $\bar{y}_{24}$ |

Note—The $\bar{y}_{ij}$ are sample means that provide unbiased estimates of the $\mu_{ij}$.

errors vary from one procedure to the next and these may be found, along with worked examples, in Marascuilo and Levin (1970).

## *Interactions as Planned Comparisons*

Some researchers have a number of misconceptions concerning the partitioning of the sum of squares in complex designs. Whereas most know that they may generate one-degree-of-freedom tests "within" a main effect that may be tested as planned orthogonal comparisons, they fail to realize that a partitioning of the interaction sum of squares is also possible when planned contrasts are specified. The prevailing belief is that while main effects may be decomposed into one-degree-of-freedom contrasts, interactions must be assessed in the context of multi-degree-of-freedom omnibus $F$ tests. That this is not true is illustrated in the following example.

Consider a cross-sectional study consisting of two factors, sex (male and female) and age (6, 8, 10, and 12 years of age) as portrayed in Table 1. Further suppose that an investigator has reason to believe that a certain cognitive ability is of such a nature that in the primary grades there is a large sex difference in favor of girls, but that the difference diminishes over the elementary school years. This statement has the flavor of an interaction hypothesis which could be evaluated on a post hoc basis following the rejection of the hypothesis of no interaction, using Scheffé's procedure based on three degrees of freedom.

But reconsider the investigator's hypothesis. On the surface, it appears that the hypothesis states that the mean profile for the boys over the four age levels is not parallel to the corresponding profile for the girls. Actually, it is more explicit than that. It states that a relatively large initial girl–boy difference will be observed that will decrease as age increases. If the investigator's hypothesis is correct, then symbolically:

$$(\mu_{11} - \mu_{21}) > (\mu_{12} - \mu_{22}) > (\mu_{13} - \mu_{23}) > (\mu_{14} - \mu_{24}).$$

In this case, it is easy to relate the interaction hypothesis to an interaction test for trend using the coefficients for linear, quadratic, and cubic components. By referring to a standard table of orthogonal polynomials, as in Hays (1963) or in Kirk (1968), one may use the same coefficients that test for trend within the main effects sum of squares to test the trend interaction hypothesis within the interaction sum of squares. The matrix in Table 2 has been constructed to incorporate the appropriate contrast coefficients.

The first column defines the contrast for determining the sex effect, that is, the difference

TABLE 2

Contrast Matrix for Table 1 Based on Tests for Linear, Quadratic, and Cubic Trends

| Cell mean | Contrast | | | | | | |
|---|---|---|---|---|---|---|---|
| | $\hat{\Psi}_S$ | $\hat{\Psi}_{A\,(\text{linear})}$ | $\hat{\Psi}_{A\,(\text{quadratic})}$ | $\hat{\Psi}_{A\,(\text{cubic})}$ | $\hat{\Psi}_{S\times A\,(\text{linear})}$ | $\hat{\Psi}_{S\times A\,(\text{quadratic})}$ | $\hat{\Psi}_{S\times A\,(\text{cubic})}$ |
| $\bar{y}_{11}$ | 1 | −3 | 1 | −1 | −3 | 1 | −1 |
| $\bar{y}_{12}$ | 1 | −1 | −1 | 3 | −1 | −1 | 3 |
| $\bar{y}_{13}$ | 1 | 1 | −1 | −3 | 1 | −1 | −3 |
| $\bar{y}_{14}$ | 1 | 3 | 1 | 1 | 3 | 1 | 1 |
| $\bar{y}_{21}$ | −1 | −3 | 1 | −1 | 3 | −1 | 1 |
| $\bar{y}_{22}$ | −1 | −1 | −1 | 3 | 1 | 1 | −3 |
| $\bar{y}_{23}$ | −1 | 1 | −1 | −3 | −1 | 1 | 3 |
| $\bar{y}_{24}$ | −1 | 3 | 1 | 1 | −3 | −1 | −1 |

between girls' and boys' performance ignoring age; the next three columns constitute three orthogonal contrasts that test for the main effect of age by means of a trend analysis for linear, quadratic, and cubic components. These coefficients are read directly from Table VI of Hays (1963). The last three columns are generated from the first four: Column 5 is found by multiplying the coefficients of Columns 1 and 2 to produce the linear Sex × Age interaction contrast, Column 6 is the product of Columns 1 and 3, while Column 7 is the product of Columns 1 and 4. These latter two sets of coefficients define the quadratic and the cubic Sex × Age interaction contrasts, respectively.

In this example, the contrast of primary interest is defined by the coefficients of Column 5. This contrast is given by:

$$\hat{\Psi}_{S\times A\,(\text{linear})} = (-3)\bar{y}_{11} + (-1)\bar{y}_{12} + (+1)\bar{y}_{13} + (+3)\bar{y}_{14} + (+3)\bar{y}_{21} + (+1)\bar{y}_{22} + (-1)\bar{y}_{23} + (-3)\bar{y}_{24}$$

which may be written as:

$$\hat{\Psi}_{S\times A\,(\text{linear})} = -\,3(\bar{y}_{11} - \bar{y}_{21}) - 1(\bar{y}_{12} - \bar{y}_{22}) + 1(\bar{y}_{13} - \bar{y}_{23}) + 3(\bar{y}_{14} - \bar{y}_{24}).$$

This is seen to have the same basic form as that used for the linear trend for main effects except that the coefficients in this case are applied to the mean sex *difference* at each of the four age levels. In addition, it should be noted that $\hat{\Psi}_{S\times A\,(\text{linear})}$ is a valid interaction contrast. In the mathematical model, $\hat{\Psi}_{S\times A\,(\text{linear})}$ is an estimate of:

$$\begin{aligned}\Psi_{S\times A\,(\text{linear})} &= -3(\mu+\alpha_1+\beta_1+\gamma_{11}-\mu-\alpha_2-\beta_1-\gamma_{21})\\ &\quad -1(\mu+\alpha_1+\beta_2+\gamma_{12}-\mu-\alpha_2-\beta_2-\gamma_{22})\\ &\quad +1(\mu+\alpha_1+\beta_3+\gamma_{13}-\mu-\alpha_2-\beta_3-\gamma_{23})\\ &\quad +3(\mu+\alpha_1+\beta_4+\gamma_{14}-\mu-\alpha_2-\beta_4-\gamma_{24})\\ &= -3(\gamma_{11}-\gamma_{12})-1(\gamma_{12}-\gamma_{22})+1(\gamma_{13}-\gamma_{23})\\ &\quad +3(\gamma_{14}-\gamma_{24}),\end{aligned}$$

which contains only interaction parameters and, thus, is not confounded with other effects.

Clearly, $\hat{\Psi}_{S\times A\,(\text{linear})}$ gets to the heart of the investigator's query, that is, whether there exists a decreasing sex difference in the cognitive ability as a function of increasing age. His question is evaluated statistically by weighting the four girl–boy differences by the appropriate coefficients that are related to linear trend. If the investigator were interested in other characteristics of the girl–boy differences, the higher order trend components could be examined individually.

It is worth mentioning that the hypothesis of mean girl–boy differences predicts that the differences will *decrease* as age increases. This is certainly in the mode of a directional hypothesis and therefore to achieve maximum statistical power, it should be analyzed as a directional (one-tailed) alternative. Since the hypothesis is related to a linear contrast, one may perform the test by use of the student $t$ distribution, by means of the test statistic:

$$t = \frac{\hat{\Psi}_{S\times A\,(\text{linear})}}{SE_{\hat{\Psi}_{S\times A(\text{linear})}}} = \frac{-3(\bar{y}_{11} - \bar{y}_{21}) - 1(\bar{y}_{12} - \bar{y}_{22}) + 1(\bar{y}_{13} - \bar{y}_{23}) + 3(\bar{y}_{14} - \bar{y}_{24})}{\sqrt{\dfrac{MS_E}{n}[(-3)^2 + (-3)^2 + (-1)^2 + (-1)^2 + (+1)^2 + (+1)^2 + (+3)^2 + (+3)^2]}}$$

where $MS_E$ represents the pooled (within cell) variance. If the investigator's claim is true, then one would expect that:

$$(\bar{y}_{11} - \bar{y}_{21}) > (\bar{y}_{12} - \bar{y}_{22}) > (\bar{y}_{13} - \bar{y}_{23}) > (\bar{y}_{14} - \bar{y}_{24}),$$

which when weighted by the above coefficients would produce a *negative* value of $t$. Thus, as a one-tailed test, the hypothesis $H_0$: $\Psi_{S\times A\,(\text{linear})} = 0$ should be rejected if the observed $t < t_v(\alpha)$, where $t_v(\alpha)$ is the critical value of $t$, based on the degrees of freedom associated with $MS_E$, at the $\alpha(100)$ percentile.

### *Summary*

The meaning of this entire discussion on Type IV errors manifested by interactions in ANOVA designs should be clear for the behavioral scientist. Significant interactions examined as either planned or post hoc comparisons must be evaluated either in terms of the interaction parameters of the model or in terms of cell means that define contrasts that reduce to comparisons among the interaction parameters of the model. If it is seen that the expected value of a contrast defined in terms of cell means contains any $\alpha$, $\beta$, or $\mu$ of the design, then it is immediately known that the contrast is not a valid interaction contrast and should therefore not be discussed as though it were related to a significant interaction component. If these principles are kept in mind and if each interesting interaction contrast is inspected in terms of its expected value, then Type IV errors in the interaction model should, like old soldiers, fade away.

## REFERENCES

Davis, D. J. Flexibility and power in comparisons among means. *Psychological Bulletin*, 1969, **71**, 441–444.

Dixon, W. J., & Massey, F. J., Jr. *Introduction to statistical analysis*. New York: McGraw-Hill, 1969.

Dunn, O. J. Multiple comparisons among means. *Journal of The American Statistical Association*, 1961, **5**, 52–64.

Dunn, O. J. Multiple comparisons using rank sums. *Technometrics*, 1964, **6**, 241–252.

Goodman, L. A. Interactions in multidimensional contingency tables. *Annals of Mathematical Statistics*, 1964, **35**, 632–646.

Hays, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.

Hummel, T. J., & Sligo, J. R. Empirical comparisons of univariate and multivariate analysis of variance procedures. *Psychological Bulletin*, 1971, **76**, 49–57.

Kirk, R. E. *Experimental design for the behavioral sciences*. Belmont, Calif.: Brooks-Cole, 1968.

Marascuilo, L. A. Large-sample multiple comparisons. *Psychological Bulletin*, 1966, **65**, 280–290.

Marascuilo, L. A., & Levin, J. R. Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. *American Educational Research Journal*, 1970, **7**, 397–421.

Marascuilo, L. A., & McSweeney, M. Nonparametric post hoc comparisons for trend. *Psychological Bulletin*, 1967, **67**, 401–412.

Miller, R. G., Jr. *Simultaneous statistical inference*. New York: McGraw-Hill, 1966.

Morrison, D. F. *Multivariate statistical methods*. New York: McGraw-Hill, 1967.

Petrinovich, L. F., & Hardyck, C. D. Error rates for multiple comparison methods: Some evidence concerning the frequency of erroneous conclusions. *Psychological Bulletin*, 1969, **71**, 43–54.

Scheffé, H. A method for judging all contrasts in the analysis of variance. *Biometrika*, 1953, **40**, 87–104.

Scheffé, H. *The analysis of variance*. New York: Wiley, 1959.

Steel, R. G. D. Some rank sum multiple comparison tests. *Biometrika*, 1961, **71**, 539–552.

(Received April 14, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 5, 375–378

# NONEXISTENT INDIVIDUAL DIFFERENCES IN REMINISCENCE[1]

EDWARD N. PETERS[2]

*Department of Pediatrics, School of Medicine and Dentistry, University of Rochester*

It is proposed that there are no individual differences in reminiscence independent of the differences in level of performance on the motor task on which the phenomenon of reminiscence is observed. Multivariate methods have failed to find such individual differences in reminiscence. The previous reports of relationships between reminiscence and organismic variables are explained as resulting from artifacts in the methods used.

Reminiscence is the gain in performance that is observed following a period of rest from massed practice on a motor task. This gain in performance has been attributed to the dissipation during rest of reactive inhibition accumulated during prior performance. Reactive inhibition has been assumed to be a variable on which individual differences exist independent of the subject's level of performance on the task. These differences in reactive inhibition have been related theoretically to a number of organismic variables such as motivation (Eysenck, 1962), introversion–extraversion (Eysenck, 1956), classroom learning (Lynn, 1960; Otto, 1965), abnormal behavior (Grassi, 1968), and retardation (Wright, 1970).

## *Methods of Measurement*

The most common method of measuring reminiscence has been a measure of performance, taken after a rest period, following massed practice minus a measure of performance, taken just prior to the rest period. This crude gain measure of the change in performance over the period of rest is the phenomenon of reminiscence and is, therefore, the most obvious measure of the inhibition dissipated. The inadequacies of crude gain measures of change and the alternatives have been documented (Manning & DuBois, 1962; Tucker, Damarin, & Messick, 1966). One alternative, the residual gain score, may be improved in the reminiscence situation by residualizing the postrest score with respect to several prerest trials, thus removing more of the prerest performance variance from the postrest measure (Peters, 1967). This approach may be extended through factor analysis.

The first factor extracted in the factor analysis of motor learning tasks of the sort that show reminiscence is usually a factor of general performance which accounts for a major portion of the total variance. Residualizing the postrest trials with respect to this factor should provide the maximum removal of variance attributable to general performance level with the data available. The inclusion of several postrest trials in the data subjected to the factor analysis would permit the extraction of a factor of that variance of the postrest trials that is independent of the factor of general performance, assuming there is such variance.

Eysenck (1960) attempted to improve the reliability of crude gain measures of reminiscence on a pursuit rotor task by measuring reminiscence under circumstances free of conditioned inhibition. He collected data on 45 subjects who took 66 trials on a pursuit rotor task with 10 periods of rest interspersed after every six massed trials. The 10 crude gain measures of reminiscence produced by these data were factor analyzed by Eysenck yielding two factors, both of which were loaded positively on some measures of reminiscence. Eysenck's (1960) conclusion from this analysis was that "the picture is not clear [p. 22]." Factor analysis of the raw scores on all trials leads to a stronger conclusion.

A principal component factor analysis with

---

[1] Research leading to this paper was carried out in part under Contract 816(14) between the Office of Naval Research and Washington University. Computer services were supported by National Science Foundation Grant G 22296 to Washington University.

[2] Requests for reprints should be sent to Edward N. Peters, Department of Pediatrics, 1545 Mt. Hope Avenue, Medical Center, University of Rochester, Rochester, New York 14620.

TABLE 1

Factor Analysis of Eysenck's Data

| Trial | Factor | | | | |
|---|---|---|---|---|---|
| | I | II | III | IV | V |
| | Eigenvalues | | | | |
| | 36.42 | 4.16 | 2.34 | 2.10 | 1.82 |
| | Percent Variance | | | | |
| | 55.18 | 6.30 | 3.54 | 3.19 | 2.76 |
| Prerest | | | | | |
| −3 | .775 | −.008 | .011 | .032 | −.001 |
| −2 | .771 | .074 | .015 | −.025 | .044 |
| −1 | .749 | .021 | .023 | .038 | .032 |
| Postrest | | | | | |
| +1 | .691 | −.002 | .090 | −.070 | −.161 |
| +2 | .755 | −.006 | −.137 | .007 | −.016 |
| +3 | .745 | −.080 | −.112 | −.017 | −.010 |

Note.—The factor loading values are the means of 10 trials.

diagonals of unity was performed on all 66 trials of Eysenck's data. The 10 rest periods were not observed to have any effect on the pattern of loadings in the factor matrix. The mean loading for the three trials before and after each rest is shown in Table 1, with the eigenvalues and the proportion of the total variance accounted for by each of the first five factors extracted. The first two factors, (*a*) a factor of general performance and (*b*) a factor of change in performance during the task, are the same ones on which previous investigators have found individual differences from data without interspersed rest periods (Jones, 1970). After these two factors, the loadings on the factors show no meaningful pattern, and the eigenvalues are too low for them to contain meaningful variance common to a number of trials.

The pursuit rotor task is not the only motor task that has been used for measuring individual differences in reminiscence. Another common task is that of inverted alphabet printing. Data on 277 subjects were collected on 18 trials at this task with two 10-minute rest periods yielding reminiscence on the ninth and eighteenth trials (Peters).[3] The factor analysis of these data yielded a first principal component factor of general performance on the task which accounted for 70% of the total variance. The residual correlation between the two postrest trials following the extraction of this factor was −.01.

Evidently, the postrest trials do not contain any common variance that is not also contained in the prerest trials on either of the two most common tasks used in research on reminiscence. Therefore, one may conclude that there are no individual differences in reminiscence independent of differences in the level of general performance on the task. This conclusion is supported by evidence for the absence of individual differences in motor performance as a function of distribution of practice (Kientzle, 1946, 1949) and effort to respond (Jones, 1968, 1969).

Why, if there are no individual differences in reminiscence independent of general performance, is there so much literature reporting evidence of the relationships between reactive inhibition as measured by reminiscence and other organismic variables? These relationships have been reported because of artifacts in the methods used to measure reactive inhibition and its relationship to other variables.

One source of artifactual relationships results from the use of crude gain measures of reminiscence. If the prerest and postrest measures have the same correlation with an organismic variable but their variances are different, then the crude gain measure will correlate spuriously with the organismic vari-

TABLE 2

Comparison between Prerest and Postrest Variance

| Rest | Variance | | $r$ | $t$ |
|---|---|---|---|---|
| | Prerest | Postrest | | |
| First | 23.75 | 30.30 | .555 | $2.43^{*}$ |
| Second | 23.86 | 30.93 | .732 | $3.17^{**}$ |

* $p < .02$.
** $p < .01$.

---

[3] E. N. Peters. *A residual gain measure of reminiscence and classroom learning.* (Office of Naval Research Tech. Rep. No. 14, Contract Nonr 816 (14)) Saint Louis, Mo.: Washington University, 1968.

able (Manning & DuBois, 1962). The variances of the prerest and postrest trials around the two rest periods on the inverted alphabet printing task (Peters, see Footnote 3) are shown in Table 2. The variances between the prerest and the postrest trials over both rest periods are significantly different at the .05 level. Thus, the metric changes from the prerest to the postrest trials on the inverted alphabet printing task. This is probably responsible for the reported correlations between reminiscence on the inverted alphabet printing task and other variables (Eysenck, 1962; Lynn, 1960; Meier, 1961).

This artifact in crude gain measures of reminiscence may also account for the correlations between separate measures of reminiscence (Eysenck, 1956). Since crude gain measures are not free of performance variance, correlations between crude gain measures may only demonstrate the extent to which the measures are contaminated with performance variance.

Not all of the studies of the relationship between reminiscence and organismic variables have been correlational studies. Another method has been to use two separate groups and to perform an analysis of variance on the scores of the prerest and postrest trials of a motor task (Otto, 1965; Otto & Fredricks, 1963). A significant interaction between groups and trials has been interpreted as evidence for a relationship between reminiscence and the organismic variable on which the groups were defined. This method is used, in part, because the groups differ on initial, prerest level of performance. The significant interactions reported indicate that the effects of trials and groups are not additive. This may be due to a difference in the metric for the two groups at the different levels of performance.

A methodologically different way of obtaining the same results is to compare the crude gain measures of reminiscence for the two groups (Grassi, 1968; Higgins & Mednick, 1963; Jones & Ellis, 1962; Shamberg, Baker, & Burns, 1969; Wasserman, 1951; Wright, 1970). However, the demonstration of significantly different amounts of reminiscence for the two groups at different levels of performance may show only that the metric is different at the different levels of performance. When performance levels are nearly equal, there may still be inhomogeneity of variance between groups. Efforts to avoid confounding of groups and levels of performance by practicing to a common level of performance (Wright & Willis, 1969) only succeed in confounding groups with practice.

What are the theoretical consequences if one accepts the null hypothesis that there are no individual differences in reminiscence independent of the differences in the level of performance on the task? One consequence is the elimination of theories relating reactive inhibition to organismic variables. The previous theoretical positions taken with regard to these relationships have proved to be neither fruitful nor wrong. Most of the support for such theories comes from research using reminiscence as the measure of reactive inhibition. One study of the relationship between reactive inhibition and introversion–extraversion that did not involve the measurement of reminiscence found no support for the theory (Yates & Laszlo, 1965). Thus, these theories seem to depend upon the existence of individual differences in reminiscence for their vitality.

## REFERENCES

Eysenck, H. J. Reminiscence, drive, and personality theory. *Journal of Abnormal and Social Psychology*, 1956, **53**, 328–333.

Eysenck, H J. Reminiscence, extraversion, and neuroticism. *Perceptual and Motor Skills*, 1960, **11**, 21–22.

Eysenck, H. J. Reminiscence, drive and personality—revision and extension of a theory. *British Journal of Social and Clinical Psychology*, 1962, **1**, 127–140.

Grassi, J. R. Performance and reminiscence of brain-damaged, behavior-disordered, and normal children on four psychomotor and perceptual tests. *Journal of Abnormal Psychology*, 1968, **73**, 492–499.

Higgins, J., & Mednick, S. A. Reminiscence and state of illness in schizophrenia. *Journal of Abnormal and Social Psychology*, 1963, **66**, 314–317.

Jones, M. B. Intertrial correlations under variations in effort. *Ergonomics*, 1968, **11**, 175–181.

Jones, M. B. Differential processes in acquisition. In E. A. Bilodeau (Ed.), *Principles of skill acquisition*. New York: Academic Press, 1969.

Jones, M. B. A two-process theory of individual differences in motor learning. *Psychological Review*, 1970, **77**, 353–360.

Jones, R. W., & Ellis, N. R. Inhibitory potential in rotary pursuit acquisition by normal and defective subjects. *Journal of Experimental Psychology*, 1962, **63**, 534–537.

Kientzle, M. J. Properties of learning curves under varied distributions of practice. *Journal of Experimental Psychology*, 1946, **36**, 187–211.

Kientzle, M. J. Ability patterns under distributed practice. *Journal of Experimental Psychology*, 1949, **39**, 532–537.

Lynn, R. Individual differences in introversion–extraversion, reactive inhibition, and reading attainment. *Journal of Educational Psychology*, 1960, **51**, 318–321.

Manning, W. H., & DuBois, P. H. Correlational methods in research on human learning. *Perceptual and Motor Skills*, 1962, **15**, 287–321.

Meier, M. J. Interrelationships among personality variables, kinesthetic figural aftereffect, and reminiscence in motor learning. *Journal of Abnormal and Social Psychology*, 1961, **63**, 87–94.

Otto, W. Inhibitory potential in good and poor achievers. *Journal of Educational Psychology*, 1965, **56**, 200–207.

Otto, W., & Fredricks, R. C. Relationship of reactive inhibition to reading skills attainment. *Journal of Educational Psychology*, 1963, **54**, 227–230.

Peters, E. N. A residual gain measure of reminiscence. In K. M. Wientge, P. H. DuBois, & H. Gaffney (Eds.), *Psychological research in classroom learning*. (School of Continuing Education Research Publication No. 11) St. Louis, Mo.: Washington University, 1967.

Shamberg, N., Baker, S., & Burns, J. Reminiscence and pursuit rotor performance in introverts, extraverts and neurotics. *British Journal of Social and Clinical Psychology*, 1969, **8**, 375–382.

Tucker, L. R., Damarin, F., & Messick, S. A base-free measure of change. *Psychometrika*, 1966, **31**, 457–473.

Wasserman, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *Journal of Experimental Psychology*, 1951, **42**, 162–172.

Wright, L. Pre-rest performance levels of nonretardates and retardates, and reminiscence on a manual crank task. *American Journal of Mental Deficiency*, 1970, **75**, 304–308.

Wright, L., & Willis, C. Reminiscence in normals and defectives. *American Journal of Mental Deficiency*, 1969, **73**, 700–702.

Yates, A. J., & Laszlo, J. I. Learning and performance of extraverts and introverts on the pursuit rotor. *Journal of Personality and Social Psychology*, 1965, **1**, 79–84.

(Received May 6, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 5, 379–385

# RETRIEVAL RULES AND SEMANTIC CODING IN SHORT-TERM MEMORY[1]

A. D. BADDELEY[2]

*University of Sussex, Brighton, England*

Studies of semantic coding in short-term memory are examined for evidence of semantic effects in the primary memory component. It is argued that all cases in which semantic factors have been shown to affect primary memory are attributable to the subject's utilization of retrieval rules, that is, techniques and strategies that are themselves stored in secondary memory but that facilitate the retrieval of phonemically coded information from primary memory. It is concluded that semantic coding produces durable secondary memory traces.

In the present paper the term short-term memory is used operationally to refer to a situation that typically involves a single presentation of a small amount of material followed within seconds by a recall test. The labile memory system that is assumed to play a prominent role in such situations is termed primary memory (Waugh & Norman, 1965). The term long-term memory refers to a situation typified by the multiple presentation of a large amount of material that may be tested after a considerable delay. The relatively stable memory system assumed to operate in such situations is referred to as secondary memory.

There is by now a considerable body of evidence to suggest that verbal short-term memory tends to involve phonemic (acoustic or articulatory) coding, and is relatively insensitive to semantic factors, while for long-term memory the reverse is true (Baddeley, 1970; Shulman, 1971). There is, however, considerably less agreement as to whether or not the two types of coding are associated with distinct types of memory, and more specifically whether semantic coding ever occurs in primary memory.

One position states that type of coding determines the durability of the memory trace, with phonemic coding of verbal material being simple and rapid and leaving an adequate primary memory trace accompanied by a relatively minor secondary memory component, whereas the semantic coding of unrelated material is relatively slow, but leaves a much more durable secondary memory trace (Baddeley & Ecob, 1970). If secondary memory is defined in terms of trace durability, then semantic coding should not occur in primary memory, since anything coded semantically will always leave a durable and hence secondary memory trace. In a recent review of similarity effects in short-term memory, Shulman (1971) cited a number of studies showing semantic effects in short-term memory, and on closer analysis concluded that semantic coding does occur in primary memory, and is not associated with slower forgetting than phonemic coding. This claim is disputed on the grounds that the results cited by Shulman and others as evidence of semantic coding in primary memory reflect the subject's use of semantically coded retrieval rules which, though themselves stored in secondary memory, are used to interpret phonemically coded primary memory traces.

## RETRIEVAL RULES

In any memory experiment, the subject remembers not only the material itself but also a number of additional "rules" that he must follow if he is to perform adequately. These include information about what he must attend to and how and when he must respond. In addition, he himself is likely to formulate "rules" that will help him perform more effectively. Thus, the subjects in digit span experiments virtually never give letters as intrusions, applying what Underwood (1964) termed a "selector mechanism" and what is called here a retrieval rule to decide, for example, that a particular deteriorated trace represents "8" and not "*h*." Such rules are clearly semantic in nature and sufficiently long term to last throughout the experiment. Unlike a retrieval cue (Tulving, 1968), which may comprise any feature of any item or sequence that will assist retrieval, a retrieval rule acts as a control process (Atkinson & Shiffrin, 1968) based on an abstract description of all or part of the material.

---

[1] This paper was written while on leave of absence from the University of Sussex, at the Center for Human Information Processing, University of California, San Diego, supported by National Institute of Mental Health Grant MH15828.

[2] Requests for reprints should be sent to A. D. Baddeley, Psychology Department, University of Stirling, Stirling FK9 4LA, Scotland.

A good example of the use of a retrieval rule is provided by a recent study by Laurence (1970) in which two paired-associate lists were learned; in one condition the second list comprised the same stimulus words as the first list, each paired with a homophone of the appropriate first list response (e.g., noise–cent; noise–scent). Over half of her subjects noticed this relationship and reported using a "same sound–other spelling strategem," which allowed perfect transfer from prior learning. Further examples of retrieval rule effects in memory experiments are given by Bower (1970). It seems likely also that given the opportunity, the subject will formulate rules applying to a particular input sequence. Thus, the digit sequence 1, 2, 3, 4, 5, 6 might be encoded as "ascending series 1–6," which would allow perfect recall without recourse to primary memory. As Bartlett (1932) pointed out, the subjects in memory experiments typically show a constant "effort after meaning"; the search for retrieval rules is one aspect of this. The next section considers the evidence for semantic coding in primary memory in the light of the retrieval rule concept. In doing so, Shulman's (1971) recent and thorough review of the topic is closely followed.

### *Semantic Similarity in Immediate Recall*

Studies cited by Shulman as evidence of semantic coding in immediate recall fall into two categories: those in which subjects attempted to recall sequences of words with similar versus different meanings and those in which sequences comprised items from more than one conceptual category (e.g., digits and letters). It has been known at least since Harden (1929) that when sequences comprise both digits and letters, presenting the items in two separate blocks (e.g., four digits followed by four consonants) will facilitate immediate recall. Such patterning effects have also been shown by Schwartz (1966), and by Sanders and Schroots (1968), and are readily interpretable in terms of the retrieval rule hypothesis, as is a comparable paired-associate study by Calfee and Peterson (1968). Provided the subject can store the information about the pattern of digits and consonants, the amount of order information required to reconstitute the sequence will be drastically reduced. In addition, the knowledge that the item in a particular position should be, for example, a digit rather than a consonant will reduce item information by allowing the subject to narrow down the number of plausible alternatives and hence increase the probability of interpreting a deteriorated phonemic trace correctly. It seems, then, that category membership effects in immediate memory are readily interpretable in terms of retrieval rules, and do not necessarily imply semantic coding in primary memory.

The observation (Baddeley, 1966) that immediate memory for sequences of words was impaired when the words were similar in meaning is not open to such an explanation. However, the effect was small, and interpretation of immediate recall effects is difficult since they are likely to reflect both primary memory and secondary memory components. In order to separate the two, it is necessary to include a delayed recall condition comprising only the secondary memory component. Using this technique, Baddeley and Ecob (1970) have shown effects of both acoustic similarity and semantic compatibility on immediate recall of word triads. After a 20-second delay, the acoustic similarity effect had virtually disappeared, while the semantic compatibility effect was undiminished, strongly indicating a secondary memory effect. A subsequent study by Baddeley and Levy (1971) attempted to produce semantic similarity effects in serial recall by using compatible adjective–noun sequences. There was no similarity effect on immediate recall, but a significant effect occurred after a 20-second delay, again suggesting that the semantic effect was associated with secondary memory and was swamped by phonemically coded primary memory effects on immediate recall.

### *Retroactive Interference*

Some of the clearest evidence of semantic coding in short-term memory comes from a study by Dale and Gregory (1966) using the Brown-Peterson task (Peterson & Peterson, 1959) and a retroactive interference paradigm. They required their subjects to read a sequence of three words that were either phonemically or semantically similar. These were recalled after a 4-second delay during which the subject read a sequence of six words that were either similar to the critical triad or dissimilar. Recall was impaired by both phonemic and semantic similarity to about the same extent. This result is commonly interpreted as evidence for both interference theory and semantic coding in primary memory. There is, however, evidence that the Brown-Peterson task reflects mainly secondary memory, having only a small primary memory component (Baddeley & Scott, 1971). Moreover, a recent study suggests that a retrieval rule interpretation is more appropriate (Baddeley, Ecob, & Scott [3]).

The retrieval rule explanation of this result is as follows. In the retroactive interference task, the subject is presented with two sets of words and subsequently required to recall the first of the two sets. When both sets are drawn from the same category, the subject is forced to rely on some form of temporal or ordinal cue or tag (Yntema & Trask, 1963). When they are drawn from mutually exclusive categories, however, category membership forms an additional cue that will allow the subject to discard the irrelevant interpolated set at retrieval, provided he has encoded this information in a suitable retrieval rule. It was predicted on the basis of this hypothesis that if mixed triads were used comprising words from three separate categories, no such simple retrieval rule would be possible; hence no retroactive interference effect should occur, although the interpolated words came from the same three categories. Interference theory would predict a retroactive interference effect in this condition, though of reduced magnitude, since the degree of overall similarity between the two sets is reduced. An experiment was therefore carried out that replicated the Dale and Gregory study in manipulating both phonemic and semantic similarity but that also studied retroactive interference effects with mixed triads. The results clearly supported the retrieval rule hypothesis by showing a retroactive interference effect for pure triads, but no trace of an effect with mixed sequences. The inclusion of mixed sequences also allowed intralist similarity effects to be estimated, given a retrieval rule interpretation. There was a marked negative effect of phonemic similarity, but no significant effect of semantic similarity, in contrast to the retrieval rule effect that is assumed to be dependent on secondary memory, and that operated equally well on the basis of either semantic or phonemic similarity.

Two further retroactive interference studies are cited by Shulman (Brown, 1958; Corman & Wickens, 1968) that showed better retention of consonants when the intervening reading tasks comprised only digits than when the task also contained consonants. These results are, of course, equally amenable to a retrieval rule explanation, since the presence of consonants in the interpolated set prevents the subject from using the simple rule "recall consonants and ignore digits" and forces him to rely on temporal cues.

### *Proactive Interference*

The most dramatic demonstrations of semantic effects in short-term memory are found using the release-from-proactive-interference technique (Wickens, 1970), which is based on the Brown-Peterson short-term forgetting task. Since release-from-proactive-interference studies typically use delays of 10 to 20 seconds, they probably contain no primary memory component. They probably represent particularly clear examples of the operation of retrieval rules in secondary memory, since conceptually the release from proactive interference phenomenon is closely analogous to the Dale and Gregory retroactive interference design, except that the subject is attempting to select the most recent set of material, and the distracting material is provided by prior sequences rather than vice-versa. Evidence for this, rather than a traditional interference theory interpretation, comes from a study by Turvey (1968), which shows that proactive interference develops just as rapidly when successive triads each comprise three items from different categories, as when all items in all triads are drawn from the same category.

---

[3] A. D. Baddeley, J. R. Ecob, & D. Scott. Retroactive interference effects in short-term memory: A cognitive interpretation with implications for the role of acoustic and semantic coding. Paper presented at the meeting of the Psychonomic Society, San Antonio, Texas, 1970.

### *Free Recall*

The free recall technique is particularly suitable for separating primary memory and secondary memory effects (Glanzer & Cunitz, 1966). Unfortunately, it is not well suited for showing phonemic coding, since the most common techniques for showing such coding rely on the tendency for phonemic similarity to disrupt the retention of order information, which is not required in free recall. However, Craik (1968) has shown that intrusion errors that seem likely to come from the primary memory portion of the list (i.e., immediately recalled terminal items) tend to be phonemic confusions that are otherwise comparatively rare. Conversely, there is no evidence of semantic coding in the primary memory component of free recall. Tulving and Patterson (1968) have shown that recall of the last four items presented in a free recall list, items that typically reflect the primary memory component, is facilitated when the four words are semantically related (e.g., north, south, east, west). However, it has been shown that the clustering effect is limited to secondary memory (Craik & Levy, 1970) and is unaffected when the primary memory component is eliminated by interpolating a filled delay (Levy & Baddeley, 1971).

### *Memory Probe Tasks*

A range of probe tasks has been used to study semantic coding in short-term memory. Of these, minimal paired-associate learning is perhaps the least suitable for studying primary memory, since Peterson (1966) has shown that only the recall of the last pair of items is influenced greatly by primary memory, and then only when tested immediately. In view of its large secondary memory component, it is perhaps surprising that so many studies have failed to find semantic similarity effects with this technique (Baddeley, 1970; Baddeley & Dale, 1966; Dale, 1967). A recent study by Baddeley and Levy (1971) suggested that subjects have difficulty in encoding unrelated words semantically under the conditions of rapid presentation typically used with this technique. When semantic encoding was facilitated by using compatible noun–adjective pairs (e.g., priest–holy, vicar–pious), a semantic similarity effect occurred, while no such effect was found when semantically unrelated pairs were used (e.g., priest–tasty, vicar–delicious).

One technique that allows a clear separation between primary and secondary memory is Waugh and Norman's (1965) sequential probe task. Using this technique, Levy and Murdock (1968) found that when lists of words were drawn from the same taxonomic category, secondary memory recall improved while the primary memory component was unaffected. Lists of phonemically similar words or consonants, on the other hand, showed impaired primary memory recall but did not affect secondary memory performance. A subsequent study by Kintsch and Buschke (1969) also used the probe technique and again found that phonemic similarity impaired primary memory but did not affect secondary memory, while semantic similarity affected secondary memory but not primary memory, although in this case the semantic effect was negative.

The studies just described present some of the clearest evidence that primary memory is not affected by semantic factors. Curiously enough, the probe technique also produces what at first appears to be the clearest evidence to the contrary, in studies by Bregman (1968) and Shulman (1970). In Bregman's study, the subject was presented with a booklet, each page of which contained either a stimulus word or a recall test. He studied each word for 3 seconds before turning the page over, and was tested after intervals, ranging from 3 to 28 seconds, that were filled by other items or tests. The tests involved prompting on the basis of rhyme, taxonomic category, or the first few letters of the word, or being given the word that had preceded the stimulus word. Thus, if "lead" were the stimulus word, the cues would be "rhymes with bread," "a chemical element," "is spelled le —," or "followed the word X" (where X was the word immediately preceding "lead"). He found comparable amounts and rates of forgetting for all except the sequential probe, for which performance was rather worse, a feature that appeared to puzzle Bregman. Before discussing this study, a similar experiment (Shulman, 1970) is described.

Shulman's study involved presenting 10-word sequences at rates of 350, 700, or 1,400 milliseconds per word. Recognition was then tested by cueing with a probe word that could be identical to a word in the stimulus set or could be a synonym or a homonym. The retention functions for the three types of probe were similar, except that the probability of correct recognition was markedly lower for the synonym probe at all parts of the forgetting curve. Response latency measures paralleled the forgetting curves with faster reaction times for more recent items. There was a pronounced difference in reaction times associated with the three probe types, with identical probes being fastest, homonym probes next, and synonym probes slowest. Shulman interpreted these reaction time effects in terms of Sternberg's (1966) concept of memory search, and pointed out that the difference between the three probes is attributable to a difference in time to encode the probe stimulus rather than in time to search the list for a matching item, since the three reaction time functions differ in intercept rather than slope. He concluded that both phonemic and semantic coding effects occur in primary memory and that they are associated with comparable rates of forgetting.

Both Shulman and Bregman's experiments differ from the previously described probe studies in testing item, but not order information. Before concluding that this is the crucial variable, however, an alternative explanation of these results should be considered. It appears to be the case from the reaction time data that the major difference between the three probe types occurs in encoding the probe word. Since the subject is told 1.3 seconds before the word appears what type of probe will be presented, it is not clear what this crucial encoding process involves. If semantic encoding takes much longer than phonemic, it is strange that (*a*) the effect of rate of presentation on synonym probe recall, though statistically reliable, is very small (correct "yes" response probability changes from .71 to .74; corrected for "guessing," these become .61 and .68); and (*b*) there should be so little effect of probe type on memory scanning rate, which presumably involves multiple semantic comparisons. There is, however, a very simple interpretation of the encoding process that fits in neatly with the rest of Shulman's results; namely, the subject attempts to recode the homonym and synonym probe word into identical probe words and then performs a phonemic memory scan. On this hypothesis, then, the subject encodes the original list phonemically; when given probe type (e.g., homonym) and probe word (e.g., sea) he generates the relevant item (see) and compares the item with his phenomically encoded primary memory trace. In the case of the identical probe no such recoding is necessary; with a homonym probe, some recoding will be required, since slight differences in pronunciation may occur between the two words. However, the probability of generating the wrong word from the probe is very low, hence the error rate is no greater for homonym than for identical probes. In the case of a synonym, the generation task is more difficult, leading to longer reaction times, and the chance of producing the wrong word is greater, leading to the higher error rate observed. It has been suggested (Shulman, personal communication, November, 1970) that the reaction times involved are too short to allow the subject to generate a synonym. If the subject were generating *ab initio,* this might well be true, but in fact the appropriate phonemic unit has been primed immediately beforehand by the presentation of the stimulus word list. The more recent the word, the greater the degree of priming and presumably the more rapidly and accurately the subject will produce the appropriate response, hence the somewhat steeper recency slope for the synonym probe observed by Shulman for both recall probability and reaction time.

The same recoding interpretation can be applied to Bregman's study; namely, when given the probe cue, the subject generates one or more words that are consistent with the cue and then decides whether that word was presented, rather than comparing each word he has experienced with each probe until he obtains a match, as Bregman's interpretation implies. With the sequential probe, no such recoding rule is possible, with the result that performance is particularly bad in this condition, which is what Bregman observed. It seems likely that both Shulman's and Bregman's results show that when the subject is given a semantic cue for material that is stored phonemically, he will solve the problem by recoding the semantic probe into a phonemic one. He is employing a retrieval rule that is *not* provided by the experimenter, but which *is* demanded by the task.

### *Recall of Sentences*

One excellent point made by Shulman is that few of the studies that have looked for semantic coding in primary memory have required the subject to attempt any form of semantic coding. Shulman did, however, exclude one important class of probe studies for which this is not true, namely, those concerned with retention of meaningful prose. The technique is best illustrated in a study by Sachs (1967), which aimed to use retention as a means of deciding between two theories of syntax. She presented sequences of prose, and occasionally tested retention of a sentence by repeating that sentence either unchanged, changed semantically, or with semantics unchanged but with various grammatical transformations. She found that subjects were unable to detect changes in syntax or word order after a very few seconds of interpolated material, whereas, even at her longest delay, semantic information showed virtually no forgetting. Johnson-Laird and Stevenson (1970) used sentences constructed on the basis of various psycholinguistic hypotheses and obtained results suggesting that only semantic content is durably encoded. This was also the conclusion reached by Bregman and Strasberg (1968) who studied recognition of sentences. An ingenious incidental learning technique was used by Wanner[4] whose prose material comprised the instructions for an experiment; the subject was required to read them and then turn over the page. This revealed a recognition test on memory for the experimental instructions. Like Sachs, he found excellent recall of semantic content, together with an even more rapid loss of nonsemantic information.

## CONCLUSION

Although it is clear that semantic coding occurs in short-term memory, there is no good evidence of semantic coding in primary memory. In all cases where this has been claimed, the results can be attributed to secondary memory. Such effects occur either because the technique used does not allow the separation of primary and secondary memory components, or else because secondary memory is used to store semantic retrieval rules that facilitate retrieval of information from a phonemically encoded primary memory trace. There is therefore no reason to abandon the simple generalization that semantic coding produces a durable memory trace.

---

[4] H. E. Wanner. On remembering, forgetting and understanding sentences. Unpublished doctoral dissertation, Harvard University, 1968.

## REFERENCES

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. II. New York: Academic Press, 1968.

BADDELEY, A. D. Short-term memory for word sequences as a function of acoustic semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 362–365.

BADDELEY, A. D. Acoustic and semantic similarity in minimal paired-associate learning. *British Journal of Psychology*, 1970, **61**, 335–343.

BADDELEY, A. D. Language habits, acoustic confusability and immediate memory for redundant letter sequences. *Psychonomic Science*, 1971, **22**, 121–122.

Baddeley, A. D., & Dale, H. C. A. The effect of semantic similarity on retroactive interference in long and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 417–420.

Baddeley, A. D., & Ecob, J. R. Simultaneous acoustic and semantic coding in short-term memory. *Nature*, 1970, **227**, 288–289.

Baddeley, A. D., & Levy, B. A. Semantic coding and short-term memory. *Journal of Experimental Psychology*, 1971, **89**, 132–136.

Baddeley, A. D., & Scott, D. Short-term forgetting in the absence of proactive interference. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 275–283.

Bartlett, F. C. *Remembering*. Cambridge: Cambridge University Press, 1932.

Bower, G. Organizational factors in memory. *Cognitive Psychology*, 1970, **1**, 18–46.

Bregman, A. S. Forgetting curves with semantic, phonetic, graphic and contiguity cues. *Journal of Experimental Psychology*, 1968, **78**, 539–546.

Bregman, A. S., & Strasberg, R. Memory for the syntactic form of sentences. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 396–403.

Brown, J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

Calfee, R. C., & Peterson, R. E. Effect of list organization on short-term probe recall. *Journal of Experimental Psychology*, 1968, **78**, 468–474.

Corman, C., & Wickens, D. D. Retroactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 16–19.

Craik, F. I. M. Types of error in free recall. *Psychonomic Science*, 1968, **10**, 353–354.

Craik, F. I. M., & Levy, B. A. Semantic and acoustic information in primary memory. *Journal of Experimental Psychology*, 1970, **86**, 77–82.

Dale, H. C. A. Semantic similarity and the A–B A–D paradigm in STM. *Psychonomic Science*, 1967, **9**, 79–80.

Dale, H. C. A., & Gregory, M. Evidence of semantic coding in short-term memory. *Psychonomic Science*, 1966, **5**, 75–76.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

Harden, L. M. The quantitative study of the similarity factor in retroactive inhibition. *Journal of Experimental Psychology*, 1929, **2**, 421–430.

Johnson-Laird, P. N., & Stevenson, R. Memory for syntax. *Nature*, 1970, **227**, 412.

Kintsch, W., & Buschke, H. Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, 1969, **80**, 403–407.

Laurence, M. W. Role of homophones in transfer learning. *Journal of Experimental Psychology*, 1970, **86**, 1–7.

Levy, B. A., & Baddeley, A. D. Recall of semantic clusters in primary memory. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 8–13.

Levy, B. A., & Murdock, B. B., Jr. The effects of delayed auditory feedback and intra-list similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 887–894.

Peterson, L. R. Short-term verbal memory and learning. *Psychological Review*, 1966, **73**, 193–207.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, **2**, 437–442.

Sanders, A. F., & Schroots, J. J. F. Cognitive categories and memory span: II. The effect of temporal vs. categorical recall. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 373–379.

Schwartz, F. Morphological coding in short-term memory. *Psychological Reports*, 1966, **18**, 487–492.

Shulman, H. G. Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 499–508.

Shulman, H. G. Similarity effects in short-term memory. *Psychological Bulletin*, 1971, **75**, 399–415.

Sternberg, S. High speed scanning in human memory. *Science*, 1966, **153**, 652–654.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Tulving, E., & Patterson, R. D. Functional units and retrieval processes in free recall. *Journal of Experimental Psychology*, 1968, **77**, 239–248.

Turvey, M. J. Analysis of augmented recall in short-term memory following a shift in connotation. *British Journal of Psychology*, 1968, **59**, 131–137.

Underwood, B. J. The representativeness of rote learning. In A. W. Melton (Ed.), *Categories of human learning*, New York: Academic Press, 1964.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

Yntema, D., & Trask, F. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 65–74.

(Received May 10, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 5, 386–402

# WHITE NORMS AND BLACK MMPIs:
## A PRESCRIPTION FOR DISCRIMINATION?[1]

MALCOLM D. GYNTHER[2]

*Saint Louis University*

A review of the literature disclosed distinctive differences between MMPIs of blacks and whites. Differences in social desirability ratings of items and disproportionate representation of black-favored items on the key scales partially account for these findings. Furthermore, education, residence, and cultural separation influence the degree of difference found. However, blacks, whether normal or institutionalized, generally obtain higher scores than whites on Scales *F*, 8, and 9. Item and factor analyses reveal that these differences represent differences between blacks and whites in values, perceptions, and expectations rather than differences in level of adjustment (as a traditional interpretation of the findings would imply). The principal value exhibited by blacks has been labeled distrust of society or social cynicism. Studies undertaken for this paper suggest that prospective black employees are disadvantaged when the MMPI is used for screening and that black psychiatric patients are less likely than whites to be diagnosed accurately by the MMPI. Various solutions to this problem are discussed; the most satisfactory approach appears to be construction of an MMPI based on black norms.

Robert Williams, speaking as National Chairman of the Association of Black Psychologists, recently called for an end of testing of black children and their parents (Williams, 1970). He was primarily concerned about intelligence, achievement, and aptitude tests that labeled blacks as uneducable, led to their placement in special classes, or assigned them to educational tracks. More generally, he and the Association of Black Psychologists opposed testing practices that would "destroy positive growth and development of Black people [Williams, 1970, p. xxviii]."

Are there problems inherent to personality assessment, specifically by means of the Minnesota Multiphasic Personality Inventory (MMPI), analogous to those Williams has underlined in the ability testing field? To answer this question and related ones, a review of the work on MMPI performance of blacks and whites is necessary. The following questions seem to summarize the critical issues: (*a*) Are there differences between MMPI performance of blacks and whites attributable to cultural factors over and above differences expected from individual variations, sex, socioeconomic status, and other influential variables? (*b*) If so, what is the nature of these differences? (*c*) If there are differences, what is the reason for them? (*d*) If differences exist, what are the consequences in terms of employment, admission to graduate or professional schools, screening psychiatric admissions for specific treatments, or other situations where the MMPI might be used? and (*e*) If differences lead to problems, as suggested above, what should be done to rectify the situation?

One might think that the answer to at least the first of these questions would be clear after more than 20 years of data collection. Such is not the case. In their most recent review of comparative psychological studies of blacks and whites in the United States, Dreger and Miller (1968) stated:

> Until other controlled studies are made to confirm or disconfirm these findings, we shall accept the conclusion that individual differences and possibly sex

---

[1] An earlier version of this paper titled "Different Cultures . . . Different Norms?" was presented at the Sixth Annual MMPI Symposium, Minneapolis, Minnesota, on April 7, 1971.

[2] The writer would like to thank Harold Altman, Philip Emmite, Marilyn Marshall, Edward Nishball, James Wallhermfechtel, and Robert Warbin for their assistance in data collection. Thanks are also extended to James Armentrout, Jeanne Miller, Milton Strauss, and Ollie Walton for reading and commenting on earlier versions of the manuscript.

Requests for reprints should be sent to Malcolm D. Gynther, Department of Psychology, Saint Louis University, Saint Louis, Missouri 63103.

differences may be more important than race differences in MMPI-revealed temperament [p. 30].

Also, Baughman (1971) in his just-published *Black Americans* asked, "how much of the discrepancy between black and white [MMPI] scores can be assigned to cultural variation? [p. 73]." He concluded, "At present, this question—referring to *degree* or *quantity*—cannot be answered [p. 73]."

The first two questions—are there MMPI differences associated with cultural differences and, if so, what kind of differences—are so interrelated that they are considered together.

## COMPARISONS OF MMPIs OF BLACKS AND WHITES

### *Studies Reviewed by Dahlstrom and Welsh*

It is relatively easy to summarize the studies carried out on blacks between 1939 and 1960. Of the approximately 1,200 studies reviewed by Dahlstrom and Welsh in *An MMPI Handbook* (1960), a grand total of 4 involved blacks. Three of these studies were of prisoners in state institutions (Caldwell, 1954; Fry, 1949; Panton, 1959); the other was of tuberculosis patients in a veterans administration hospital (Hokanson & Calden, 1960). The first study was that of Fry (1949) and, in retrospect, his finding of no significant Scale score differences between black and white inmates of two northern prisons probably tended to diminish whatever potential interest there may have been in this topic. In any case, 5 years passed before another article comparing blacks and whites was published. In this case, differences *were* reported, notably that blacks scored substantially higher than whites on Scale 8 (*Sc*). The two other studies available to Dahlstrom and Welsh at the time of their review obtained significantly higher Scale scores by blacks on Scales 8 and 9 (*Ma*). The writers of the handbook interpreted these results as "the sort that would be expected from known effects of socioeconomic inequities [p. 273]." In other words, the repeated finding of differences on Scales 8 and 9 should not be attributed to cultural differences between blacks and whites, but rather to social class differences between blacks and whites. Dahlstrom and Welsh also pointed out that none of these studies provided samples that would be considered representative of normal blacks. Obviously being a tuberculosis patient or prisoner might interact with one's cultural background in unknown ways to influence MMPI performance. Before leaving what might be called the "early work" on the effects of different cultural training on MMPI performance, which was inconclusive due to the samples used and the failure to control for social status, it might be of interest to mention that Harrison and Kass (1967) reanalyzed Fry's data and found that his black subjects *did* obtain significantly higher Scale 9 scores than his white subjects. So these four studies agreed that blacks obtain higher scores than whites on Scales 8 and/or 9, even if the case for differences associated with subcultures had to be regarded as "not proved."

### *More Recent Findings*

The first published study of normal blacks was conducted by Ball (1960). His sample was composed of ninth-grade students in two central Kentucky towns and included only 31 blacks, 14 boys and 17 girls. He compared these subjects with larger numbers of white boys and girls who attended the same classes as the black students. He found that the black males obtained significantly higher Scale 1 (*Hs*) scores than their white counterparts and that black girls obtained significantly higher scores than white girls on Scales *F*, 8, and 0 (*Si*) and significantly lower scores on Scale 3 (*Hy*). High-point analysis revealed no significant differences between the male samples, but significant differences among the females, apparently due to a relatively high frequency of Scale 4 (*Pd*), 9 peaks among the white girls and a relatively high frequency of Scale 6 (*Pa*), 8 peaks among the black girls. He noted that his blacks were predominantly from low-status families, while the whites were mostly from higher status families.

McDonald and Gynther (1962) then published a similar study, but with much larger samples of blacks (i.e., over 10 times the number that Ball had used). Their subjects were high school seniors from segregated schools in and around Charleston, South

Carolina. The white subjects were principally middle or upper class, the blacks mainly lower class. To avoid undue repetition and detail, only those scales on which black males and females differed significantly from the white subjects are reported. The black subjects were significantly higher than the whites on Scales *L*, *F*, 1, 2(*D*), and 9. Comparison of high-point codes classified by the Haertzen-Hill procedure revealed no striking differences between the different subgroups, although there was a tendency for the black males to be overrepresented in the category that included Scales 1, 2, 3, and 7 (*Pt*) and the white males to be overrepresented in the category that included Scales 5 (*Mf*) and 0.

The first study to use normal subjects and control for social status was published by McDonald and Gynther in 1963. This investigation might be considered the first reasonably adequate test of the proposition that MMPI differences between blacks and whites are attributable to cultural differences. A factorial design was used with sex, race, and social status being the factors controlled. High school seniors attending segregated schools served as subjects. Results consistent with Dahlstrom and Welsh's explanation of the pre-1960 work were possible; that is, marked scale score differences associated with class, with few or no differences associated with race, could have been found. But that did not occur; no differences associated with the class factor were found on *any* scale. Furthermore, differences associated with race were as striking as on any previous study with class not controlled. Blacks got significantly higher scores than whites on Scales *L*, *F*, 2, 5, 8, and 9.

The study just cited yielded clear-cut results, but confirmation with other samples was still essential. Butcher, Ball, and Ray (1964) soon published a comparable study using college students. They found that blacks obtained significantly higher scores than whites on the *L* scale and that whites obtained significantly higher scores on Scale 6 than blacks, with social status controlled. Other differences associated with race were found, but these were dependent also on sex of the subjects. These results are less impressive than those of McDonald and Gynther, possibly because Butcher's samples contained fewer than one-fourth the subjects used by McDonald and Gynther. Nevertheless, Butcher and his associates concluded that "subcultural differences exist independent of sex and socio-economic level [p. 86]."

The next major study was carried out by Harrison and Kass (1967). Although their samples were of women only, they are very large and make use of more adequate sampling techniques than in the previously reported studies, which used largely "captive audiences." Subjects were successive admissions of pregnant women to a city hospital prenatal clinic, which was on a nonpayment basis with no private care. Differences between social status of the black and white groups were not formally measured, but since all subjects came from the underprivileged, lower-class neighborhoods surrounding Boston City Hospital, these differences cannot have been very great. This supposition is supported by the finding of no significant difference on the *K* scale, which has been shown to be related to social status. Scale score analysis disclosed that blacks obtained significantly higher scores than whites on Scales *F*, 1, 8, and 9. Harrison and Kass also subjected their data to item analysis with remarkable results: 213 of the MMPI items discriminated between blacks and whites at the .05 level or better! Anyone who has used this method for comparing groups will realize that this magnitude of difference has seldom, if ever, been found in MMPI studies.

Harrison and Kass factor analyzed the 150 most discriminating items. Those findings are considered in the next section, which focuses on possible explanations for differences between blacks and whites on the MMPI. However, Harrison and Kass' (1967) comments on scale versus item analysis seem appropriate here. They stated:

> The scales are not very sensitive to race differences, whereas the items are remarkably sensitive. A cancel-ing-out process must be at work in each scale. If so, each scale should have approximately equal number of (significant) Negro-favored and white-favored items. This is indeed the case, as is seen in Table 10. Table 10 also reveals that 32% of the discriminating items appear on none of the clinical scales [p. 462].

In more recent work, Baughman and Dahlstrom (1968) had rural North Carolina eighth graders take MMPIs and found highly significant scale score and configurational differences between the races. Black subjects were significantly higher than white subjects on Scales *F*, 2, 6, and 8. Scale 8 was shown to be substantially more likely to be the highest elevation among the black sample as compared to the white sample (e.g., 45% of black males peaked on Scale 8 versus 13% of the white males).

Another study focused more precisely on the effects of residence. Erdberg, whose unpublished dissertation was completed in 1969, used a factorial design with sex, race, and residence as the factors controlled. Social status was not explicitly controlled, but occupation, education, and income were recorded for each subject. Inspection of these data suggests that the subjects were low or middle class with minor, if any, differences among the groups. With regard to the 3 validity scales and 10 clinical scales, there was a significant main effect for race on 9 of the scales, for residence on 7 of the scales, and for sex on 3 of the scales. Blacks scored significantly higher than whites on Scales *F*, 4, 6, 8, and 9, and whites scored higher than blacks on Scales 3 and 5 (and *L* and *K*). Residence also exerted a surprisingly powerful effect on the scales; these differences reflect the fact that urban subjects obtained generally higher scale scores than rural subjects. Sex × Race and Sex × Residence interaction effects were negligible, but six significant Race × Residence interactions were reported. The direction of the differences was consistent, leading to the conclusion that black–white personality differences are less among city dwellers than among those who live in rural areas. Erdberg followed up these scale score analyses with item and factor analyses. While the outcome of the factor analyses is discussed in the next section, it can be reported that Erdberg's item analyses revealed results strikingly similar to those of Harrison and Kass; that is, he found 177 items which discriminated between blacks and whites at the .05 level or better. (Miller, Knapp, & Daniels [1968] obtained comparable results via item analyses of MMPI responses of black and white male patients at a Veterans Administration mental hygiene clinic. These results appear less critical than those of Harrison and Kass and Erdberg [1969] because focusing on normals yields "cleaner" findings, which are also more representative of blacks in general.)

What kind of results would be obtained if one were to deliberately select black subjects to be as different from white norms as possible? Rural residents should be used as subjects, preferably with no college training, with marked cultural isolation from whites, if possible, and of low social status. (Harrison and Kass quoted Pettigrew as stating that middle-class blacks attempt to minimize differences between themselves and their white middle-class counterparts, whereas these pressures are not so great among lower-class Negroes.) Gynther, Fowler, and Erdberg (1971) recently located a sample meeting these criteria. The mean profile of these 88 subjects peaked on Scale 8, with the second highest clinical scale peak on Scale 9. The average scores were approximately 80 and 75, respectively. Of the subjects 44% peaked on Scale 8, 19% on Scale 9, and 15% on Scale 6. Perhaps even more remarkable is the fact that 78 out of the 88 profiles had at least one clinical scale score of 70 or greater (73 of 88 if Scale 5 is excluded). These individuals were functioning normally according to their peers and had no history of time spent in mental hospitals or penitentiaries. One might claim that these subjects did not understand the items, but most had 8 or more years of education, and the tape-recorded version of the MMPI was used to minimize literacy differences.

Responses to selected items were also examined. These subjects' reaction to "Evil spirits possess me at times" is illustrative of these findings. This item was endorsed by less than 3% of the normative group, 14 out of 541. In this sample of rural blacks, 58% endorsed the item!

### *Summary of Research Findings*

The data reported above may be summarized as follows: (*a*) There are MMPI differences associated with differences in cultural background; (*b*) these differences most frequently appear on Scales 8 and 9, which are members of the psychotic tetrad; (*c*) item analyses disclose even greater differences associated with subculture than scale score or configurational analyses; and (*d*) the degree of MMPI differences between blacks and whites appears to be affected by such variables as education, residence, and cultural separation. (As the focus in this paper is on normal blacks, some of the post-1960 studies that used prisoners or psychiatric patients as subjects have been omitted; cf. Costello, Tiffany & Gier, 1972; Flanagan & Lewis, 1969; Miller, Wertz, & Counts, 1961.)

## EXPLANATIONS OF BLACK–WHITE MMPI DIFFERENCES

There are several possibilities that may account for the striking differences noted in the previous section: (*a*) Blacks are more maladjusted, more disturbed, more inclined to fantasy and daydreaming than whites; (*b*) blacks have different values, interests, and expectations than whites reflecting the differences in subcultures; (*c*) differences are a function of scale composition; (*d*) differences in responding are a function of differences in social desirability of the items; and (*e*) key words or phrases may have different connotative meanings to blacks and whites and hence may arouse different responses.

### *Differential Maladjustment Rates*

Caldwell (1954) obtained a mean Scale 8 score of 68 for his sample of 34 youthful black "rural-farm offenders" and said that this finding "discloses specific tendencies toward schizophrenia [p. 297]." Since the mean IQ of the group was 72 with 7 years of education on the average, it is not surprising that the Scale 8 scores were elevated. No external evidence of schizophrenia or personality disorganization was given or even looked for.

Consider Ball's (1960) interpretation of his results. He concluded:

> A comparison of Negro and white ninth grade high school students with the MMPI revealed a high incidence of neurotic tendencies among the Negro boys as well as withdrawal and introversion among the Negro girls [p. 306].

Concerning the "neurotic tendencies" phrase, his evidence is that 7 of the 14 black males peaked on Scales 1, 2, 3, or 7. At what elevation? The only significant scale score difference between the black and white boys was on Scale 1 where the means were 56 and 50, respectively. It is doubtful that many would find Ball's conclusions acceptable, let alone convincing.

Nevertheless, this interpretation cannot be rejected out of hand, especially since blacks do consistently obtain higher scale score values than whites, especially on Scales 8 and 9. Dahlstrom and Welsh (1960) stated: "high scores . . . increase the probability that subjects answering that way are members of the pathological group . . . [p. 281]"; and, further, "Clinicians will probably continue to equate overall elevation of the profile with degree of severity . . . [of emotional disorder] [p. 283]." Dahlstrom and Welsh also pointed out:

> peak scores on scale 8 are quite rare in normal adult males and females, but are much more likely in younger subjects and prison inmates. There is also an increase of these peaks in psychiatric samples, particularly hospitalized groups . . . [p. 202].

Peak scores on Scale 9, on the other hand, are relatively common, especially among younger subjects; however, 8-9 is not included in Dahlstrom and Welsh's summary of major configurational patterns.

In any case, the MMPI interpreter faced with a profile, often with a score or two above 70, and with 8-9, 8-6, 9-8, 8-4 or similar two-point peaks, is apt to consider that the individual is emotionally disturbed, possibly with schizoid mentation. An interesting example is of this orientation and its consequences is given in the next section. For now, however, let us turn from the clinical approach and ask how actuarial or computer programs might respond to 8-9 profiles in terms of blind analysis. For Marks and Seeman (1963), 70% with this profile were diagnosed as psychotic, with delusions, hallucinations, and inappropriate affect appearing as frequent symptoms. For Gilberstadt and Duker (1965), the favored diagnosis was schizophrenic reaction, catatonic type with hyperactivity, paranoid delusions, tension, and other severe symptomatology. Fowler's (1967) system would be apt to conclude that the person's test pattern resembles that of schizophrenics and to emphasize similar symptomatology as that mentioned for the other systems.

Whether the analysis is clinical or actuarial, the relatively high profiles with positive slope generated by blacks are apt to be labeled as maladjusted. There is, however, *no* external validating criteria to support this conclusion. For whites, a mass of such data has been accumulated: The friends and acquaintances' ratings used by Hathaway and Meehl,[3] the professional assessment staff's evaluations (Block & Bailey;[4] Gough, McKee, & Yandell [5]) used at the Institute for Personality Assessment and Research, Black's (1953) use of college roommates as judges, Halbower's (1955) technique of having therapists *Q* sort their patients, and the present actuarial schemes developed by Marks and Seeman (1963), Sines (1966), and others. No similar work has been published giving behavioral correlates of high-point 2s, low 6s, 4-9s, or 8-6s, or any other configuration that may occur in MMPIs given by blacks. (Baughman and Dahlstrom [1968] made a general comparison of their MMPI data with interview data and teachers' ratings. Some consistencies were found, but typically the MMPI indicated the presence of more severe personality problems than the other investigative techniques did, especially among the Negro boys. Baughman and Dahlstrom suggested that the MMPI data are identifying a latent condition that may go unrecognized as long as the child can avoid more complex environments or intensified demands.) There is some work in progress, however, that bears more directly on this issue. At the Missouri Institute of Psychiatry in Saint Louis, the writer and his associates are developing a new computer interpretation system for MMPIs of psychiatric patients. Code types are compared to the rest of the cases using a 111-item mental status inventory completed by psychiatric residents as the external criterion. In one case, a striking description of white patients was obtained; that is, there were numerous significant differences on the mental status items and the resulting picture was clinically meaningful. Application of the same analysis to our black patients yielded only 8 items out of the 111 significant at the .10 level, that is, results that are essentially meaningless.[6] In another case, code type 4-9/9-4, the same number of discriminating items was obtained for blacks as for whites. However, item-by-item comparison of the significant descriptors showed few agreements. In fact, statistical analysis indicated that the degree of agreement did not exceed what would be expected by chance. The implication of these preliminary findings is that one cannot assume that a profile or code type has the same behavioral referents for blacks and whites. The attribution of maladjustment to blacks' MMPIs on the basis of white norms appears unwarranted and highly undesirable.

### *Different Cultures, Different Values*

McDonald and Gynther stated in 1963 that "Negroes' higher scores on *L*, *F* and the various clinical scales are a consequence of a different pattern of interests, values and expectations [p. 115–116)]," but this was simply a post hoc explanation, as Dreger and Miller pointed out in their 1968 review article. McDonald and Gynther also hypothesized that the greater the separation between the races, the greater the differences to be found on MMPI scale scores. A single study has not been designed as yet to test this hypothesis, but the greatest differences found so far were obtained in the rural, isolated sample collected by Gynther et al. (1971), whereas the smallest differences appear to have been found in Butcher's (1964) study of college students. This latter result might be considered as consistent with the hypothesis. These subjects attended segregated colleges and might not have had much opportunity for contacts, but they shared values common to college students. That is, enrollment in college implies putting a certain degree of importance on learning, on striving to get ahead, and on willingness to depart from the family culture and join a youth culture. Also, one might say that blacks who attend college are more likely to accept white values and expectations than blacks who do not go to college.

---

[3] S. R. Hathaway & P. E. Meehl. Adjective check list correlates of MMPI scores. Unpublished materials, 1952.

[4] J. Block & D. Bailey. Q-sort item analyses of a number of MMPI scales. (Technical Memorandum OERL-TM-55-7) Berkeley, Calif.: Officer Education Research Laboratory, 1955.

[5] H. G. Gough, M. G. McKee, & R. J. Yandell. Adjective check list analyses of a number of selected psychometric and assessment variables. (Technical Memorandum OERL-TM-55-10) Berkeley, Calif.: Officer Education Research Laboratory, 1955.

[6] Replication of these findings has since been carried out on independent samples. For whites, replication substantially confirmed the original findings, but for blacks, not one descriptor was replicated. Moreover, half of the items found to be significant in the original sample of blacks had reversed proportions in the replication sample. For further details, see Gynther, Altman, and Warbin (in press).

This evidence of cultural effects on MMPI performance pales in significance compared to more recent, more direct findings. Harrison and Kass (1967) found over 200 MMPI items that significantly discriminated between the races. One-third of these race-sensitive items were found among those one-third of the total MMPI items that are not on any of the clinical scales. Since many of these items are related in manifest content to topics having no obvious relevance to adjustment, the implication is that the differences found are not due to differences in adjustment. If they were, one would have anticipated that a disproportionately large number of the race-sensitive items would be members of the clinical scales. The factor analysis of the race-sensitive items supports the "cultural values" position. The discriminant analysis of factor scores reveals that the four most important factors are estrangement and impulse-ridden fantasy (both higher for blacks) and self-consciousness and dislike of school (both higher for whites). These dimensions, as well as other important ones such as religiosity and romantic interest, primarily reflect attitudes and values, not psychopathology.

Harrison and Kass' (1968) second analysis of their data also has important implications for attempts to explain black–white personality differences. In this case, comparisons were made between MMPIs of blacks born in the South, blacks born in the North, and whites born in the North. If personality differences are biologically determined, the length of the black groups' exposure to a northern white culture should be unrelated to scores on the differentiating personality scales. If the differences are primarily cultural in origin, blacks who have been exposed to the northern white environment for a long time should be more similar to northern whites than those exposed for only a short time. The results clearly supported the acculturation hypothesis in that the northern black group is halfway between the southern black group and the white group on most of the personality dimensions considered. Harrison and Kass noted that there are personality differences between their northern black and southern black pregnant females; the latter admit alienated, angry, anxious feelings as well as romantic and religious interests, whereas the northern Negro woman is primarily cynical, defensive, and wary. The authors concluded that growing up in a northern urban environment appears to be emotionally constricting for a black person.

Baughman and Dahlstrom (1968) selected five of the factor scales developed by Harrison and Kass to provide additional data on how their 13- and 14-year-old youngsters viewed themselves. They found the most dependable differences on scales labeled cynicism and intellectual and cultural interests. Blacks were substantially higher on both measures. Blacks were also significantly higher on the estrangement scale. Baughman and Dahlstrom (1968) concluded:

> In the area of emotional ties and interpersonal relationships, these children show pervasive mistrust of themselves and others, extreme pessimism about receiving fair treatment or justice, and expectations that people will be self-seeking, dishonest, and double dealing [pp. 256–257].

Erdberg (1969) also factor analyzed his race-sensitive items. The most important factor was labeled "distantiation" and was defined as feelings of cynicism about the motives of others. Blacks were higher than whites. This factor would seem to be tapping the same dimension as Harrison and Kass' "cynicism." The other named factors were absence of optimism, admission of fears, unusual thought patterns, and aspiration–reality conflict. Blacks were higher on all these factors. This group seems to reflect both values and attitudes and possibly psychopathology. Concerning the possibility that more psychopathology among the blacks is suggested by the greater admission of fears and unusual thoughts, Erdberg remarked:

> There does not seem to be good evidence for . . . more actual psychopathology for Negroes in this sample. All subjects were functioning community members and there was no behavioral evidence . . . [of psychopathology] [p. 63].

Erdberg gave two other explanations to account for his results:

> A less defensive, more literally honest response set for Negroes, as evidenced by (significantly) lower scores on *L* and *K* or a different set of cultural norms for Negroes which may make experiences such as hearing voices acceptable in some circumstances [p. 62].

In any case, the major factor found by Erdberg is "social cynicism," certainly an important and realistic value in the black culture.

Miller et al. (1968) did not factor analyze their race-sensitive items, but did group these items subjectively into rationally meaningful clusters. Their subjects, it should be remembered, were Veterans Administration mental hygiene clinic patients. The clusters that they isolated were labeled physical symptoms, strong incorporation of conventional standards, denial of interpersonal anxiety, and mistrust of society. The clusters involve complaints about general health, endorsement of moralistic values, denial of anxiety in social situations, and mistrust of societal institutions and the motives of individuals. Blacks scored higher on every cluster. A careful comparison of these results with those of Harrison and Kass shows much overlap between the content areas isolated in the two studies.

A common theme that runs through all of these studies is the marked distrust of society exhibited by blacks. Grier and Cobbs (1969), the black psychiatrists who recently published *Black Rage,* focused on precisely this trait. They stated:

> We submit that it is necessary for a black man in America to develop a profound distrust of his white fellow citizens. For his own survival, he must develop a *cultural paranoia* in which every white man is a potential enemy unless proved otherwise and every social system is set against him unless he personally finds out differently [pp. 149–150].

These writers view this as an "adaptive trait developed in response to a peculiar environment [pp. 149–150]" and feel that it is "no more pathological than the compulsive manner in which a diver checks his equipment before a dive or a pilot his parachute [pp. 149–150]."

### *Differences Associated with Scale Composition*

Are there any intrinsic features of the scales themselves that could lead to differences? One could point out that Harris and Lingoes'[7] first *Sc* subscale is labeled "social alienation," which is very similar to the first factor derived by Harrison and Kass and by Erdberg. Also, it might be noted that the *F* scale is composed of items rarely endorsed by *whites;* what items might emerge from a similar analysis of responses by blacks is not known. However, some data presented by Harrison and Kass (1967) appear especially pertinent. Scales *K*, 2, or 7 seldom differentiate between blacks and whites. The number of significant items favored by blacks versus the number of significant items favored by whites on those scales are 4 to 5, 10 to 9, and 7 to 7, respectively; that is, there is an equivalent number of race-sensitive items on each of these scales, with little discrimination associated with differences in cultural background to be expected. Scales *F*, 8, and 9 are those on which blacks most frequently score higher than whites. Black–white endorsement on these scales is 14 to 4, 19 to 5, and 12 to

---

[7] R. E. Harris & J. C. Lingoes. Subscales for the MMPI: An aid to profile interpretation. San Francisco: Department of Psychiatry, University of California, 1955. (Mimeo)

5, respectively. Scale 3 is the only scale on which whites frequently obtain higher scores than blacks. The ratio is reversed; 7 of the 20 race-sensitive items are favored by blacks, 13 by whites. Confirmation of these relationships would seem to call for obtaining social desirability scale values from blacks and whites, especially on the race-sensitive items. Apparently blacks answer items true or false partly at least in terms of how they feel the blacks they know would approve or disapprove the behaviors or sentiments involved. There are no published social desirability or favorability ratings of MMPI items by blacks, so the writer obtained such ratings to clarify this issue. Before reporting these results, however, it seems worth noting what the two most significant race-sensitive items on Scales *F* and 8 are. For Scale 8, the two most significant items seem to reflect the black experience. The items are "I believe I am a condemned person" and "People say insulting and vulgar things about me," both answered "true" significantly more often by blacks. For Scale *F*, the two most discriminating items are what many would regard as typical black endorsement of conservative religious beliefs: "A minister can cure disease by praying and putting his hand on your head" and "I am very religious (more than most people)."

### *Differences in Social Desirability Values*

Are these massive differences in responding to MMPI items, reported by Harrison and Kass and other writers, a function of black–white differences in social desirability ratings of these items? If this relationship is extremely powerful, one might expect that social desirability ratings of nearly all of the race-sensitive items would differ significantly between blacks and whites. One would also predict that the correlation between blacks' percent endorsement of the items and their social desirability values would be very high, possibly between .85 and .90 as Edwards (1964) has typically reported for the groups he has investigated.

If Harrison and Kass had obtained social desirability ratings, as well as completed MMPIs, or if the writer had been able to obtain endorsements of MMPI items in addition to social desirability ratings, we would be in an ideal position to check out the relationships just posited. Those data are unfortunately not available. However, social desirability ratings of the 144 most discriminating race-sensitive items were obtained from 50 black and 50 white adults in Saint Louis, and Harrison supplied the true–false endorsement rates of these items by his large Boston sample.

A brief description of these 100 subjects is in order. Our pool of subjects, both black and white, included teachers, nurses, cooks, maids, laborers, custodians, social workers, insurance underwriters, housewives, hair stylists, accountants, and others. About two-thirds of the 100 subjects were women. The average age of the blacks was 41 years, as compared to 39 for the white sample. The average number of years of education was 13.1 for blacks, 14.9 for the whites. Analysis of social class by means of Hollingshead and Redlich's (1958) index placed the average white near the bottom of Class II and the average black near the top of Class III. The match of these Saint Louis samples is fairly good on these dimensions, but the subjects are clearly of higher social class than the Boston sample.

Of the 144 comparisons, there were 48 differences between social desirability ratings of the black and white subjects significant at the .05 level, with 29 of these 48 significant at the .01 level or better. The differences between means of blacks and whites were in the predicted direction in three-fourths of the discriminating items. An example may clarify how social desirability results were predicted. Of Harrison and Kass' black subjects 31.8% answered "true" to the item "I am sure I am being talked about"; 17.3% of the whites gave the same response. The social desirability ratings of this item among the blacks in the Saint Louis sample should be more favorable (and they were). More impressive results would probably have been obtained if lower class or rural subjects had been used, but even these findings suggest strongly that social desirability ratings of MMPI items by blacks and whites are apt to be quite different.

The data permit two other comparisons, namely, the correlation between the Saint

Louis blacks' social desirability ratings and the percentage "true" endorsement by Harrison and Kass' black sample and the correlation between social desirability ratings of the Saint Louis white subjects and the percentage "true" endorsement of Harrison and Kass' white sample. The correlations are based on mean values on these variables for the 144 items. The correlation for black subjects was .67, suggesting a moderate relationship between social desirability values and true–false endorsement of items. This value is substantially lower than Edwards' usual finding and raises the question of whether blacks are less influenced in their responses to MMPI items by social desirability than whites. The alternative explanation of this comparatively low correlation involves our use of two groups of subjects from different regions and social classes instead of the usual one-group analysis. The degree of correlation for whites provides a solution to this problem; if the correlation is in the .80s, the first alternative is probable; if in the .60s, the second alternative is likely to be correct. The correlation for white subjects was .63. The appropriate conclusion appears to be that although blacks and whites differ as to their perception of the items, both groups are approximately equally influenced by the social desirability variable.

### *Differences in Connotative Meaning of MMPI Statements*

The fifth possibility, that many of the words and phrases that make up the items have different connotative meanings for blacks than for whites, has not received much attention by MMPI researchers. (Dahlstrom [personal communication, March 1971] commented that "A student . . . found very little support for the notion that the differences [in the Orange county data] were arising from vastly different interpretations of either content material or adverbial modifiers.") However, cross-cultural research with the Semantic Differential Scale (Snider & Osgood, 1969) indicates that such differences exist. Williams (1966), for example, found that blacks and whites responded very differently to color names, and ethnic concepts, especially on the evaluative (good–bad) factor. Blacks rated the concept "brown person" more positively than either "black person" or "white person," while whites rated such concepts as "white," "white person," and "Caucasian" most "good" and the concepts "black," "black person," and "Negro" as most "bad." With regard to the MMPI, it would appear that blacks and whites might respond very differently to such concepts as "soul" or "policeman."

### *Summary: Explanations of Black–White MMPI Differences*

The following generalizations of the data reviewed in this section are offered: (*a*) Differences in MMPI performance of blacks and whites reflect differences in values, perceptions, and expectations that result from growing up in different cultures; (*b*) the most striking value exhibited by blacks is distrust of white society (alternate terms might be social cynicism or alienation); (*c*) there is no satisfactory evidence to indicate that blacks' MMPI performance should be interpreted as revealing that they are less well adjusted than whites; (*d*) intrinsic features of the MMPI, that is, scale composition, partially account for scale score and configural differences; and (*e*) differences in social desirability values of the items and, possibly, differences in connotative meanings of certain words, account in part for black–white differences in responding.

## PRACTICAL CONSEQUENCES OF BLACK–WHITE MMPI DIFFERENCES

Despite the hue and cry of several years ago about "invasion of privacy," MMPIs are still used in many places to screen applicants for employment or admission to some program. They are also used routinely, of course, to help diagnose patients in mental hospitals or clinics. And some universities give them to all entering freshmen.

The writer has not discovered any published studies relating to the issue of differential consequences for blacks and whites on the basis of MMPIs given in the situations just indicated. If differential consequences could be demonstrated, especially differential negative consequences, this would be a serious matter that would require the attention of those authorities concerned with discriminatory practices. In fact, the United States Supreme Court ruled in March 1971 that tests which are discriminatory cannot be used to make decisions with regard to hiring or promotions.

In the absence of published work, the writer has attempted to put together some data on this matter. Some of it is only suggestive; other portions of it are more directly relevant.

### *Applicants for a Police Force*

First, a brief analysis of MMPIs of black and white applicants for positions as policemen on a large metropolitan force is given. It should be noted that the MMPI is given in the third step of the screening process, so that differences would tend to be minimized by the earlier screening. First, the applicant must be between 21 and 31 years of age, at least 5 feet 7 inches tall, have no physical handicaps, possess a driver's license, never have been convicted of a felony, be a high school graduate or hold a Graduate Equivalency Diploma and have an honorable discharge from the service. Second, the applicant must score above a certain level on the Police Aptitude Test (a measure of intelligence), be fingerprinted, have these prints sent to the FBI, have his police record checked, and submit a notarized application. At this point, he is given the MMPI.

Subjects were 100 white and 100 black individuals who were applicants in 1966, 1967, and 1968. The major findings were that 16 of the 100 white applicants were ultimately rejected, and 15 of the 100 white applicants' MMPIs had at least one clinical scale score of 70 or above. There was little congruence between these two events. In fact, only 4 of those rejected were among the 15 with the relatively high profiles. Of the 100 black applicants 27 were rejected and 23 of the 100 blacks had MMPIs containing at least one clinical scale score of 70 or above. Again, congruence was poor. Only 7 of the 27 rejected had relatively high-ranging MMPI profiles. Two conclusions seem warranted: (*a*) More blacks had high clinical scale scores than whites, as would have been expected from the earlier review of the literature (here it is especially noteworthy, as the applicants who took the MMPI had to be similar in so many ways); and (*b*) more blacks than whites were ultimately eliminated from consideration.

Other relationships found between code types and rejection or acceptance are not reported because it is unclear what role, if any, the MMPI played in the decision-making process. The MMPI results were not typically communicated to those in charge of hiring; on the other hand, it is clear that the MMPIs were not simply filed and ignored. The individuals responsible for this phase of the screening say that on "infrequent" occasions they recommended referral to a psychiatrist on the basis of the MMPI profile. They are unable to specify what criteria they followed. This may be an example of differential consequences associated with differential MMPI performance, but it cannot be proved.

### *Applicants for Hospital Attendant Positions*

A second approach to this problem involved individuals who were applicants for work as attendants in a public mental hospital in the Saint Louis area. The vast majority of these applicants are black, and hence a white comparison group is not available. Perhaps the white applicants for the police force could be considered as a comparison group, but it would not be very adequate in view of the differences in hiring procedures and considering what would presumably be different motivations for seeking to be a policeman versus a hospital attendant. The MMPI is administered to attendant applicants more or less as they walk in off the street. Other evaluative procedures including interviews are, of course, carried on, but the applicant takes the MMPI the same day. The only screening, then, involves getting an Otis score, which translates into an IQ of 80 or better.

This procedure has been carried on at this installation since 1958. To obtain an adequate sample which was also recent, the files for 1967, 1968, and 1969 were pulled. The total number of subjects was 258, of which approximately two-thirds were males. In this case, the number ultimately accepted or rejected is unknown, since that decision is made by the nursing staff. However, the number recommended for acceptance or rejection by the psychologists is available. The psychologists during the period were what might be described as "MA plus" psychologists with a considerable degree of experience, but only a fair amount of training in MMPI interpretation.

The results for the total sample were as follows: In 1967, 35% were designated as not acceptable; in 1968, 38% were categorized as not acceptable; and in 1969, 35% were classified as not acceptable.

At that point, approximately every third case was drawn from these alphabetical files for a closer evaluation of what determined acceptance or rejection. This sample of 96 closely resembled the complete files for these years in that 35 were described as not acceptable. An equivalent percentage of males and females were rejected, so these data were combined.

Where psychologists are functioning as MMPI evaluators, it is clear that profile elevation plays a major role in their decision making. Of these 96 applicants, 31 had at least one clinical scale score at 70 or above; of these 31, 24 were rejected or described as very questionable (which is functionally equivalent to a recommendation to reject). Of the 96 applicants, 65 had no clinical scale scores of 70 or above; 54 of these individuals were described as acceptable. These differences are highly significant.

The data also included "remarks," especially when the evaluation was negative. Of the 11 cases who were rejected with no scores of 70 or above, three had low Scale 9 scores, which were interpreted as "low energy, difficult to motivate"; 2 had high *L* scores, raising questions as to "sincerity and insight"; 4 had Scale 4 scores in the upper 60s which were interpreted as suggesting "difficulty in relating to authority and accepting supervision"; 1 had a scale 6 score in the 60s; another had a Scale 6 score in the low 40s. "Overly sensitive" and "insensitive" were the respective descriptors.

To put this in context, we are dealing with blacks who are trying to get jobs as hospital attendants. It is clear that those who are rejected by the psychologists who interpret the MMPI are by and large those who get relatively high scale scores, that is, at least one clinical score of 70 or greater. Our review has shown that blacks typically get higher scale scores than whites for a variety of reasons, among which differential maladjustment rate is relatively minor, if it contributes at all. However, a vicious circle has been demonstrated with the consequence that blacks will often not be recommended for employment on the basis of MMPI performance. In this instance, relatively precise discriminations were made on the basis of moderately high or low elevations on certain scales. These psychologists are not necessarily incompetent MMPI interpreters. However, they apparently do not realize that statements that might be valid for whites have not been demonstrated to hold for blacks. As their clientele are nearly all blacks, their interpretations cannot be considered racially discriminatory per se. Nevertheless, they are clearly damaging to those *individuals* who fail to get hired for this reason.

### *Diagnosis of Psychiatric Patients*

As a third example of possible real-life consequences of differential MMPI performance of blacks and whites, preliminary results of a study of psychiatric patients that Milton Strauss and the writer are doing is reported. The question here is whether black and white patients might not have differential misclassification rates; that is, will MMPI diagnoses correspond as well to final psychiatric diagnoses for blacks as they will for whites? The assumption is that these two diagnoses will not agree as frequently for blacks as for whites, because of the finding that normal and institutionalized blacks routinely obtain higher scale scores, especially on Scales *F*, 8, and 9, with no clear-cut implication of greater psychopathology. In other words, there is more noise in the system when dealing with blacks' MMPIs, which should result in less exact prediction. This is a situation in which misses can have serious consequences. Diagnosis, broadly speaking, determines attitudes of staff toward the patient, type of treatment instituted, how long the patient is likely to stay in the hospital, and other important real-life matters.

The MMPIs were obtained from psychology department files of an acute intensive psychiatric treatment center which serves an indigent, racially mixed, urban population. The sample consisted of psychotics and behavior disorders. With the whites, $N$ was sufficient to draw unbiased samples of 50 psychotic males, 50 psychotic females, 50 male behavior disorders, and 50 female behavior disorders from the alphabetical files. With the blacks, only the male behavior disorders were found in sufficient quantities to use the same procedure. The $N$s for the other three black categories consisted of every case in the files. Thus, there were 50 male behavior disorders, 17 female character disorders, 33 psychotic males, and 29 psychotic females. Totals were 200 whites and 129 blacks. Age and years of education were also recorded for each patient.

With regard to making diagnostic statements from MMPIs, several courses were open. Such statements could have been taken from the reports written by the psychologists. However, at least two problems would be present using that procedure: (*a*) Some MMPI analyses would give personality descriptions, but no diagnostic summary; and (*b*) the level of training and experience of the interpreters would be extremely variable. A second possibility was for the writer to classify all profiles using a clinical impressionistic approach. However, it has been shown repeatedly that actuarial procedures are more reliable than clinical procedures. A third possibility would make use of psychotic scales or cutting scores (cf. Gynther & Petzel, 1967). Yet this approach is rather far removed from how a psychologist would usually make decisions about MMPI profiles. A fourth possibility would make use of a simple high point analysis. That is, 8s and 6s would be classified as psychotics; 4s and 9s as behavior disorders. This is workable, but does not really account for configuration. It was decided that the best approach would be to make use of the profile types developed by Marks and Seeman (1963), by Gilberstadt and Duker (1965) and by Sines (1966). In those cases in which their findings had nothing to say about the profile, we would try to match it with their closest profile type or use our best judgment, making certain that the same judgment was applied in all cases.

To give a few examples, Marks and Seeman and Gilberstadt and Duker agreed that an 8-9 indicates psychosis and that a 4-9 indicates behavior disorder. When 8-9s and 4-9s were found, that is how they were classified. Also, Sines and Gilberstadt and Duker agreed that a 4-3 suggests behavior disorder; Marks and Seeman said 2-8/8-2 and Gilberstadt and Duker said 8-1-2-3 are indicative of psychosis. It is difficult to generalize, but, in the published code types, psychosis usually involves Scale 8 in combination with other scales, and behavior disorder usually involves Scale 4 in combination with other scales (except Scale 8). The profiles were coded (to the first three high points) and classified as psychotic, behavior disorder, or neurotic without knowledge of whether the cases in question were black or white, psychotic or behavior disorder.

The results were as follows: there were 118 hits out of 200, or 59% correct classification for the white patients. For the black patients, there were 58 hits out of 129, or 45% correct classification. Statistical analysis of these findings shows that the whites were classified correctly by the MMPI significantly more often than the blacks were. The hypothesis is therefore confirmed. The implication is that, were the MMPI used to diagnose black and white patients, blacks would not receive as adequate treatment as whites because their problems and symptoms would be misunderstood more frequently. These differences, incidentally, cannot be attributed to age or education as there were no significant differences between the groups on those variables.

Further examination of these data revealed that white psychotics and character disorders were both classified more accurately by MMPIs than black psychotics and character disorders. However, the discrepancy is much more striking among the psychotics; only 34% of the black psychotics were correctly classified, compared to 56% of the white psychotics. This result is surprising. The writer would have predicted that black character disorders would be least adequately diagnosed by the MMPI on the grounds that blacks' typically higher scores on Scales 8 and 9 would lead interpreters to call such individuals "psychotic" more frequently than the psychiatric diagnosis would warrant.[8]

### *Summary: Consequences of Black–White MMPI Differences*

There is some evidence to suggest that black job applicants are disadvantaged when the MMPI is used for screening. Also, it appears that black psychiatric patients are less likely to receive optimum hospital care when the MMPI is an important part of the diagnostic battery. These studies are not elegantly designed or completely analyzed. Further work is necessary to confirm or disconfirm these tentative findings. Yet these results are consistent with expectations garnered from the review of the literature.

## APPROPRIATE AND INAPPROPRIATE SOLUTIONS TO THE PROBLEM

How can we solve the problems that follow from giving MMPIs to blacks? Clearly, the difficulties are complex and range from methodological considerations to real-life issues.

A number of approaches are possible. One solution might follow Williams' (1970) suggestion about intelligence and achievement tests; namely, stop giving MMPIs to blacks. This approach seems self-defeating, inasmuch as more, not less, needs to be known of the personalities, values, and needs of the different cultural groups in the United States. A modification of this proposal merits some consideration. Gynther et al. (1971) recently stated, concerning their results on rural, isolated black subjects: "Classification errors of this magnitude suggest the need for a moratorium on interpreting profiles generated by subjects dissimilar in background and experience from normative groups [p. 237]." Collection of MMPI data need not be halted, but interpretative activities might well be curtailed or discontinued temporarily until some empirical relationships have been demonstrated.

A second approach is to continue with business as usual. This might be described as a "head in the sand" approach. There are MMPI articles and, indeed, important MMPI books which in no place indicate whether their subjects are black or white. This omission implies that these individuals feel that their findings apply equally well to blacks or whites or, to put it another way, that MMPI differences associated with subcultural differences are so unimportant that one need not worry about them. Some people are acutely aware of the importance of interpreting MMPIs within the context of the appropriate reference group. But there are undoubtedly many instances where Gilberstadt and Duker's findings, based on Veterans Administration males, are being applied to non-Veterans-Administration females; where Marks and Seeman's results, derived from white patients at the University of Kansas Medical Center, are being applied to Spanish-Americans in Los Angeles; or where actuarial analyses, based on hospitalized psychiatric patients, are being used with college students. The same error was illustrated earlier in dealing with the black applicants for hospital attendant jobs. Perhaps the psychologists there were aware of the problem, but went ahead anyway, hoping that interpretive errors would be relatively inconsequential.

A third approach would involve making interpretive adjustments to account for known differences. Let us say, for example, that everyone agrees that blacks obtain significantly higher scores on Scales 8 and 9 than whites and, further, that these differences do not necessarily indicate that blacks are less well adjusted than whites. Then, if asked to interpret a black's MMPI which has rather prominent scores on Scales 8 and 9, say in the 70–80 range, this profile might be treated as one given by a white person with scores on these scales in the 60–70 range. This procedure is probably used by many people in interpreting profiles given by teenagers since nearly all MMPI users know that teenagers obtain higher clinical scale scores than adults. In any case, a set of rules could be developed to modify standard interpretations when one is dealing with an MMPI filled out by a black.

---

[8] Discriminant function analysis has since been applied to these data. The hit rate for black behavior disorders was significantly lower than that for the derivation or cross-validation sample of whites. The equation called too many blacks psychotic. The Goldberg (1965) index was also used. This index successfully discriminated white male psychotics from white male behavior disorders, but did not work for black males. The reason for failure in this case was that too many black male behavior disorders were called psychotic. It is interesting to note that these more objective analyses yielded results consistent with expectations, while the more subjective analysis reported above yielded inconsistent results. All analysis, however, support the major prediction, namely, that MMPI diagnosis of whites is more accurate than MMPI diagnosis of blacks.

A fourth approach would involve restandardization of the MMPI using an appropriate fraction of blacks, 11% or whatever the current United States population figure is. This is the approach that was used by Wechsler in developing the Wechsler Adult Intelligence Scale norms in response to criticism of the lack of blacks in the Wechsler-Bellevue Form I development. Obviously this would involve a great deal of work. What would it accomplish? It is quite clear that what would result are very minor shifts in the means and standard deviations of the MMPI scales. If, for example, whites had a *T* score mean of 50 on Scale 9 and blacks had a *T* score mean of 60 on the same scale, using the procedure described would result in a new mean of approximately 51. A shift of one point here or there is not going to solve this problem anymore than Wechsler's solution solved the problem of assessing the intelligence of blacks. One could also argue that since at least some of the items mean different things to blacks and whites, such a procedure would be like averaging apples and pears.

A fifth approach would involve classifying blacks' MMPIs into code types and then determining behavioral correlates for the types available. Some preliminary results of this type of analysis were reported earlier in this paper. This approach can lead to valid interpretations of MMPI profiles obtained from blacks, but would be restricted to primarily institutionalized samples. In other words, the behavioral correlates are typically derived from case histories, mental status inventories, or *Q* sorts by therapists. Similar data on normals, especially people in general, would be extremely difficult to obtain.

A sixth approach is to construct an MMPI for blacks. One could start from scratch, collecting items, etc., but a more modest (though still intimidatingly difficult) approach would involve using the current set of MMPI items, as well as the current set of scale labels. What would differ from the present MMPI is the item composition of the scales. As far as the development of clinical scales is concerned, the same procedure used in the 1940s would still seem appropriate, namely, comparison of response patterns of carefully diagnosed psychiatric patients with response patterns of normally functioning adults. In this case, however, the normal adults and psychiatric patients would be black. The validity scales could also be developed in a similar manner as in the original work. An intriguing beginning would involve developing an *F* scale for black subjects, using the same 10% or less endorsement criterion for selection of such items. If many whites got elevated *F* scores with this new scale, would they be considered as "unconventional" or "nonconforming"? Most MMPI workers would probably feel ill at ease with such an interpretation. By the same token, use of the current *F* scale with blacks is essentially meaningless. However one views this point, it is clear that a deviant score by a black on an *F* scale developed on blacks would permit us to say that this individual does not conform within the context of the black culture. Also, an elevation on Scale 8 that is based on black norms would permit us to evaluate the individual's standing among his peers with reference to whatever Scale 8 measures. This approach is probably the soundest of all those described, especially for dealing with noninstitutionalized subjects. It is possible that some of the current standard scales could be retained (i.e., those on which black–white differences have seldom been found). However, Scales *F*, 1, 8, and 9 at a minimum need to be reconstructed using black samples.

## CONCLUDING REMARKS

The evidence indicates that MMPI differences between blacks and whites are associated with differences in cultural background. The types of differences found might lead the unwary interpreter to infer that blacks are more disturbed and maladjusted than whites. However, there are no data to support this conclusion. The differences appear to be due primarily to differences in values and interests between members of the white and black subcultures. Secondary reasons such as residence, scale composition, and response set also play a role. Unfortunately, MMPI differences can disadvantage blacks, both as to employment opportunities and optimum diagnosis and treatment. Stopping the administration of MMPIs to blacks, restandardization of the test with a small percentage of blacks, or continuing as we have for so many years do not appear to be viable solutions. Alerting individuals to the seriousness of the problem and using qualifying rules for interpretation of profiles given by blacks may help temporarily, but derivation of behavioral correlates of profile types and construction of an MMPI form based on black norms are the only satisfactory solutions.

Although the focus in this paper has been on the MMPI exclusively, many of the issues considered are relevant to the use of other personality assessment devices, both objective and projective in nature. Developers and users of these other tests appear to have ignored these issues to the same extent that persons involved with the MMPI have. Clearly, these problems can no longer be neglected by individuals who are concerned with personality assessment.

## REFERENCES

Ball, J. C. Comparison of MMPI profile differences among Negro-white adolescents. *Journal of Clinical Psychology*, 1960, **16**, 304–307.

Baughman, E. E. *Black Americans: A psychological analysis*. New York: Academic Press, 1971.

Baughman, E. E., & Dahlstrom, W. G. *Negro and white children: A psychological study in the rural south*. New York: Academic Press, 1968.

Black, J. D. The interpretation of MMPI profiles of college women. *Dissertation Abstracts*, 1953, **13**, 870–871.

Butcher, J. N., Ball, B., & Ray, E. Effects of socioeconomic level on MMPI differences in Negro-white college students. *Journal of Counseling Psychology*, 1964, **11**, 83–87.

Caldwell, M. G. Case analysis method for the personality study of offenders. *Journal of Criminal Law, Criminology and Police Science*, 1954, **45**, 291–298.

Costello, R. M., Tiffany, D. W., & Gier, R. H. Methodological issues and racial (black–white) comparisons on the MMPI. *Journal of Consulting and Clinical Psychology*, 1972, **38**, 161–168.

Dahlstrom, W. G., & Welsh, G. S. *An MMPI handbook*. Minneapolis: University of Minnesota Press, 1960.

Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and whites in the United States. *Psychological Bulletin Monograph Supplement*, 1968, **70**(3), Pt. 2.

Edwards, A. L. Social desirability and performance on the MMPI. *Psychometrika*, 1964, **29**, 295–308.

Erdberg, S. P. MMPI differences associated with sex, race, and residence in a southern sample. Unpublished doctoral dissertation, University of Alabama, 1969.

Flanagan, J. C., & Lewis, G. Comparison of Negro and White lower class men on the general aptitude test battery and the Minnesota Multiphasic Personality Inventory. *Journal of Social Psychology*, 1969, **78**, 289–291.

Fowler, R. D. Computer interpretation of personality tests: The automated psychologist. *Comprehensive Psychiatry*, 1967, **8**, 455–467.

Fry, F. D. A study of the personality traits of college students and of state prison inmates as measured by the MMPI. *Journal of Psychology*, 1949, **28**, 439–449.

Gilberstadt, H., & Duker, J. *A handbook for clinical and actuarial MMPI interpretation*. Philadelphia: Saunders, 1965.

Goldberg, L. R. Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs*, 1965, **79**(9, Whole No. 602).

Grier, W. H., & Cobbs, P. M. *Black rage*. New York: Basic Books, 1968. (Republished: New York: Bantam, 1969)

Gynther, M. D., Fowler, R. D., & Erdberg, P. False positives galore: The application of standard MMPI criteria to a rural, isolated, Negro sample. *Journal of Clinical Psychology*, 1971, **27**, 234–237.

Gynther, M. D., & Petzel, T. P. Differential endorsement of MMPI *F* scale items by psychotics and behavior disorders. *Journal of Clinical Psychology*, 1967, **23**, 185–188.

Gynther, M. D., Altman, H., & Warbin, R. W. The interpretation of uninterpretable MMPI profiles. *Journal of Consulting and Clinical Psychology*, in press.

Halbower, C. C. A comparison of actuarial versus clinical prediction to classes discriminated by the MMPI. *Dissertation Abstracts*, 1955, **15**, 1115.

Harrison, R. H., & Kass, E. H. Differences between Negro and white pregnant women on the MMPI. *Journal of Consulting Psychology*, 1967, **31**, 454–463.

Harrison, R. H., & Kass, E. H. MMPI correlates of Negro acculturation in a northern city. *Journal of Personality and Social Psychology*, 1968, **10**, 262–270.

Hokanson, J. E., & Calden, G. Negro-white differences on the MMPI. *Journal of Clinical Psychology*, 1960, **16**, 32–33.

Hollingshead, A. B., & Redlich, F. C. *Social class and mental illness*. New York: Wiley, 1958.

Marks, P. A., & Seeman, W. *The actuarial description of abnormal personality: An atlas for use with the MMPI*. Baltimore: Williams and Wilkins, 1963.

McDonald, R. L., & Gynther, M. D. MMPI norms for southern adolescent Negroes. *Journal of Social Psychology*, 1962, **58**, 277–282.

McDonald, R. L., & Gynther, M. D. MMPI differences associated with sex, race and class in two adolescent samples. *Journal of Consulting Psychology*, 1963, **27**, 112–116.

Miller, C., Knapp, S. C., & Daniels, C. W. MMPI study of Negro mental hygiene clinic patients. *Journal of Abnormal Psychology*, 1968, **73**, 168–173.

Miller, C., Wertz, C., & Counts, S. Racial differences on the MMPI. *Journal of Clinical Psychology*, 1961, **17**, 159–160.

Panton, J. H. Inmate personality differences related to recidivism, age and race as measured by the MMPI. *Journal of Correctional Psychology*, 1959, 4, 28–35.

Sines, J. O. Actuarial methods in personality assessment. In B. A. Maher (Ed.), *Progress in experimental personality research*. New York: Academic Press, 1966.

Snider, J. G., & Osgood, C. E. (Eds.) *Semantic differential technique: A sourcebook*. Chicago: Aldine, 1969.

Williams, J. E. Connotations of racial concepts and color names. *Journal of Personality and Social Psychology*, 1966, 3, 531–540.

Williams, R. Report to the APA council of representatives. *American Psychologist*, 1970, 25(5), xxvii–xxviii.

(Received May 19, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 5, 403–405

# NOTE ON "FACTORS INFLUENCING THE OUTCOME OF PSYCHOTHERAPY"

H. J. EYSENCK[1]

*Institute of Psychiatry, University of London*

It is argued that it is improper to draw conclusions from studies using psychotherapy regarding causal influences unless control groups are included in the analysis (or are available in the literature) which give information about the changes that would have taken place without therapy (spontaneous remission). It is suggested that the conclusions drawn by Luborsky, Chandler, Auerbach, Cohen, and Bachrach from their 1971 survey of empirical study do not follow from the data, but may be simply artifacts due to their failure to include a proper base line in their computations.

In an important recent paper, Luborsky, Chandler, Auerbach, Cohen, and Bachrach (1971) have undertaken to review the evidence regarding factors influencing the outcome of psychotherapy. Their stated aims were threefold: to offer guidelines to clinicians; to compare clinical and quantitative factors, in the hope of indicating what should be investigated by quantitative research; and to provide a methodological evaluation of the research as a guide to future investigation. Their results are interesting and may prove in due course to have highlighted the most relevant and important factors. In this brief note, an attempt is made to take issue with a major methodological point, neglect of which may render their endeavor nugatory. Put briefly, it is suggested that when we are dealing with an attempt to study the influence of certain factors on certain types of activity, then we must include in our evaluaion control groups not exposed to any of the factors under investigation, or we must demonstrate unambiguously that in the absence of certain factors studied, there are no significant changes in the activity in question. Such a base line is absent from the review given by Luborsky et al. (1971), and there does not appear any realization in the paper that without it the conclusions drawn do not follow in any rigorous way from the data.

This would not be so serious if control groups not exposed to psychotherapy, as defined by Luborsky et al. (1971), did not in fact show significant change during time intervals comparable to those used in the studies reviewed by them. This is not so. Eysenck (1952, 1960, 1969) has reviewed the evidence on "spontaneous remission" and has shown that while the studies in this field are methodologically no more sophisticated than are those concerned with the effects of psychotherapy, nevertheless, they do show an impressive degree of agreement. Severely ill psychoneurotic patients, not treated by psychotherapy or any other type of psychiatric treatment, improve and get better to a very significant extent; the formula $X = 100\ 1 - 10^{-.00435N})$ was suggested as fitting observed results quite closely ($X =$ proportional improvement; $N =$ number of weeks since beginning of experiment). Eysenck (1960, 1969) suggested that unless treatment can be shown to improve on this no-treatment base line, then it is scientifically impossible to defend the hypothesis that treatment has in fact had any effect; all the effects observed could be due to spontaneous remission. Eysenck (1963) has also elaborated a theory of spontaneous remission that may suggest reasons for the occurrence of these curious and important effects; this theory is derived from Pavlovian and Hullian formulations of conditioning and extinction phenomena.

Kiesler (1966) has thrown doubt on the existence of spontaneous remission, and referred to it as a "myth" which, he believes, "has been perpetuated primarily by Eysenck [p. 114]." He suggested that the evidence

---

[1] Requests for reprints should be sent to H. J. Eysenck, Institute of Psychiatry, University of London, Bethlehem Royal Hospital, London, S. E. 5, England.

> seems to lead unequivocally to the conclusion that there is no evidence for spontaneous remission of psychoneurosis. Hence the belief seems to be nothing more than a myth propagated by a popularized and naive interpretation of two research studies [the studies of Landis, 1937, and Denker, 1947] [p. 119].

This statement is not correct; many more studies demonstrating spontaneous remission are quoted by Eysenck (1960), and even by Kiesler himself. Luborsky et al. may not agree with Eysenck's estimate of the importance of spontaneous remission, but if so it would have been useful had they indicated in their paper their reasons for such disagreement. The evidence, such as it is, is remarkably unanimous in its results (Rachman, 1971).

Luborsky et al. (1971) do, in fact, mention the possibility of spontaneous remission, but apparently without realizing that this factor, if left uncontrolled, may vitiate all their conclusions. They stated:

> Our survey is limited to the factors which influence change as a result of psychotherapy *only for those who have started psychotherapy. For most of the studies surveyed, therefore, it cannot be determined whether the type of individual who profits most would also have profited from another form of treatment, or from change-inducing experiences which are usually not designated as psychotherapy—or indeed from nothing more than the mysterious changes attributed to the passage of time* [p. 161].

Consider one of their main conclusions, and one which has received the most powerful support from a large number of studies.

> In 20 out of 22 studies of essentially time-unlimited treatment, *the length of treatment was positively related to outcome; the longer the duration of treatment or the more sessions, the better the outcome!* It is a temptation to conclude—and it may be an aaccurate conclusion—that if psychotherapy is a good thing, then the more the better [p. 154].

Luborsky et al. (1971) also discussed two alternative possibilities, but they made no mention of the most obvious hypothesis to explain their results: If untreated neurotics improve roughly in accordance with the Eysenck formula, and if treatment has no effect whatever, then we would expect precisely the observed effects—improvement over time! Methodologically it seems impossible to conclude anything from the studies surveyed unless it can be shown that the rate of progress was significantly faster than would be suggested by the Eysenck formula. The conclusion seems forced upon us that the neglect of spontaneous remission, and the absence of proper control groups not receiving treatment by psychotherapy, make it impossible to regard the conclusions drawn by Luborsky et al. (1971) as following rigorously from their data.

Little can be said by way of conclusion, other than that the findings of the Luborsky et al. (1971) article cannot be accepted as delineating factors influencing the outcome of psychotherapy; it is possible that the conclusions are in fact correct, but the fundamental methodological error underlying all their work makes it impossible to derive any conclusions from the facts presented. Luborsky et al. (1971) did not discuss the writer's attempts to unify the whole field by suggesting similar mechanisms at work in behavior therapy, psychotherapy, and spontaneous remission, nor did they quote papers supporting the conditioning theory of neurosis implicit in this attempt (Eysenck & Rachman, 1965); this makes it difficult to know how they would react to these arguments. It is not essential, however, for these additional hypotheses to be correct in order to see that the paradigm adopted by Luborsky et al. (1971) is in fact erroneous; all attempts to judge the importance of factors that lead to a departure from a base line require the establishment of such a base line. Without base line, no comparisions of a worthwhile, rigorous nature are possible (Eysenck, 1972).

## REFERENCES

Eysenck, H. J. The effect of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 1952, **16**, 319–324.

Eysenck, H. J. The effects of psychotherapy. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. London: Pitmans, 1960.

Eysenck, H. J. Behavior therapy, spontaneous remission and transference in neurotics. *American Journal of Psychiatry*, 1963, **119**, 867–871.

Eysenck, H. J. *The effects of psychotherapy*. New York: Science House, 1969.

Eysenck, H. J. New approaches to mental illness: The failure of a tradition. In H. Gottesfeld (Ed.), *Controversies in community health*. New York: Behavioral Publications, 1972, in press.

Eysenck, H. J., & Rachman, S. *Causes and cures of neurosis.* San Diego: R. R. Knapp, 1965.

Kiesler, D. J. Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin,* 1966, **65**, 110–136.

Luborsky, L., Chandler, M., Auerbach, A. H., Cohen, J., & Bachrach, H. M. Factors influencing the outcome of psychotherapy: A review of quantitative research. *Psychological Bulletin,* 1971, **75**, 145–185.

Rachman, S. *The effects of psychotherapy.* Oxford: Pergamon Press, 1971, in press.

(Received May 27, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 5, 406–408

# ANOTHER REPLY TO EYSENCK

LESTER LUBORSKY[1]

*School of Medicine, University of Pennsylvania*

The present article disputes Eysenck's attempt to show that the results of the 1971 survey of Luborsky, Chandler, Auerbach, Cohen, and Bachrach "may be simply artifacts." Basic disagreements are pointed out, particularly with Eysenck's reliance upon "spontaneous remission" rates, and his neglect of controlled psychotherapy studies, and of the evidence that not only patient factors but also therapist and treatment factors have an impact on the outcome.

Eysenck suggests that the conclusions of the review by Luborsky, Chandler, Auerbach, Cohen, and Bachrach (1971) "may be simply artifacts due to their failure to include a proper base line in their computations." These are strong words. They would seem to demolish the usefulness of our survey of 166 studies of the factors influencing the outcome of psychotherapy. Yet, when translated, they show he is only restating an obvious possibility that he has reiterated for many years—the *quantitative* evidence that psychotherapy helps patients is insufficient. So, he reasons, we *may* have wasted our time, since what happens to those who received psychotherapy *may* turn out to be what happens to those who did not receive psychotherapy.

In our review we have, in fact, intentionally culled those factors shown to be related to outcome of treated patients only. We explicitly and purposely did not compare the factors in treated versus nontreated patients. We stressed the possibility that some, or even all, of the factors we found might also be true for untreated patients—i.e., that some of the factors in treated patients are nonspecific. Some examples from our review of patient factors that may have such nonspecific components are "adequacy of personality functioning" and "social achievements." People who are well endowed with these qualities probably can manage more successfully to get the help they need from psychotherapy *or from other change-facilitating experiences.* We maintain, however, the value of having reviewed the factors in treated patients as an important first step, because there is evidence that psychotherapy benefits some patients more than others and because some of our predictive factors will be shown to be partially specific to psychotherapy.

This leads to the basic disagreement between Eysenck and myself—we evaluate the evidence differently. Eysenck looks at the research and decides it is weak, and it must imply that psychotherapy is not beneficial. I look at the same research, but decide it implies that psychotherapy does help people. In Eysenck's view, the main evidence (referred to in his preceding note) is derived from studies reporting rates of improvement for untreated groups ("spontaneous remission rates"), which then are used to compare with the rates from treated groups from other studies. In my view, this evidence may be leading Eysenck to draw an erroneous conclusion. The studies are inadequate as comparison groups, since there is no basis for trusting their comparability with the treatment groups. For all anyone knows, the nontreatment groups contain more patients with high "adequacy of personality functioning" and other favorable qualities for improvement, with or without treatment. Different kinds of people, by self-selection or referral, may be coming for psychotherapy versus remaining untreated. It is well known that patients who are given different forms of treatment are markedly different initially—for example, patients given psychoanalysis at the Menninger Foundation, and undoubtedly elsewhere as well, are significantly higher initially in their health–sickness ratings than patients given other forms of psychotherapy or a brief period of counseling and then discharged (Luborsky, 1962.) Obviously, the improvement rate of these patients might differ because their initial state was markedly different. An even more encompassing case has been made by Bergin (1971, p. 246) that untreated groups, which are sometimes used to establish base lines, may not really be controls at all. It should be noted here that even within the confines of utilizing the spontaneous remission studies, there is serious divergence in his use of those data—Bergin (Bergin, 1971, p. 241) after a comprehensive analysis, concluded that the median "spontaneous remission rate" appears to be only about 30%, which sharply contrasts with Eysenck's estimate.

---

[1] Requests for reprints should be sent to Lester Luborsky, University of Pennsylvania, School of Medicine, University Hospital, 207 Piersol Building, 3400 Spruce Street, Philadelphia, Pennsylvania 19104.

Eysenck does not refer to those studies that are more adequately controlled than his spontaneous remission groups. Some progress has been made in this area of research in the last two decades, since more of these controlled studies are now available than at the time of my earlier reply (Luborsky, 1954) to Eysenck's (1952) review. Among these are:

1. Frank, Gliedman, Imber, Stone, and Nash (1959). Improvement rates were compared for three matched groups of patients: once-a-week psychotherapy, group therapy, and one-half hour every 2 weeks. The group in most intensive treatment did better (i.e., once-a-week psychotherapy).

2. Nichols and Beck (1960) went to considerable lengths to try to match each psychotherapy-treated patient with a nonclinic control case. They found that *both* therapy and control groups showed significant improvement; however, the therapy group had significantly greater improvement on two crucial factors.

3. Morton (1955) compared brief psychotherapy for severe personal problems of 40 students with a matched control group. The matching was shown to be very effective. The patients were reevaluated and retested: 93% of the treated subjects and 47% of the controls showed some improvement.

4. Arbuckle and Boy (1961) did a controlled study of 36 students with major behavior problems treated by client-centered therapy. The students had been divided into three groups, carefully matched. The group receiving therapy was significantly helped in terms of better attitude toward themselves (self–ideal correlations). Teachers' ratings of behavior improved significantly for the treated, but not for the controls.

5. Shelley and Johnson (1961) reported on the beneficial effects of a 6-month counseling program in a camp for youthful offenders. The counseled group showed significantly greater decrease in antisocial themes and there was a direct relation between success in parole and reduction in antisocial themes (on the Thematic Apperception Test).

The above are brief samples of better controlled studies than those cited by Eysenck. Those, as well as other such adequate studies, are listed by Meltzoff and Kornreich (1970), who also list some controlled studies with null and negative results. They concluded, however, that

> controlled research has been notably successful in demonstrating significantly more behavioral changes in treated patients than in untreated controls. In general, the better the quality of the research, the more positive the results obtained [p. 177].

Finally, a correction that must further limit the scope of Eysenck's criticism: our review is not confined to factors in the patient, but includes factors in the therapist and in the treatment. Obviously, influences on the outcome deriving from therapist and treatment are less subject to the "spontaneous remission" argument than are the patient factors. Our review lists several studies demonstrating that some types of therapists, for example, those with more empathy, produce better outcomes with their patients. It is hardly to be expected that these empathic therapists happen to be blessed with more patients who remit spontaneously! Of the 10 studies we reviewed 8 show that certain basic similarities (e.g., in values and attitudes) between patient and therapist provide a significant advantage for the outcome of psychotherapy. Could such patient–therapist pairs also be lucky in their spontaneous remission rate?

So, for the following reasons, the results of Luborsky et al. 1971 will not be made "nugatory": (*a*) As an impetus to more focused research, it is important to systematically examine the large literature on the factors influencing the outcome of psychotherapy, even if some of these will eventually be found to be nonspecific to psychotherapy. (*b*) The "evidence" that Eysenck uses to conclude that psychotherapy is of no special value is derived from poorly controlled base line studies and is unsuited to the use he makes of them. He does not list the better controlled research, which suggests definite benefits from psychotherapy. (*c*) Eysenck's critique of the results of our review could at best (or at worst) only apply to factors in the patient, not to those related to the therapist and treatment.

## REFERENCES

Arbuckle, D., & Boy, A. Client-centered therapy and counseling students with behavior problems. *Journal of Counseling Psychology*, 1961, **8**, 136–139.

Bergin, A. The evaluation of therapeutic outcomes. In A. Bergin & S. Garfield (Eds.), *Handbook of psychotherapy and behavior change*. New York: Wiley, 1971.

Eysenck, H. J. The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 1952, **16**, 319–324.

Frank, J., Gliedman, L., Imber, S., Stone, A., & Nash, E. Patients' expectancies and relearning as factors determining improvement in psychotherapy. *American Journal of Psychiatry*, 1959, **115**, 961–968.

Luborsky, L. A note on Eysenck's article, "The effects of psychotherapy: an evaluation." *British Journal of Psychology*, 1954, **45**, 129–131.

Luborsky, L. Clinicians' judgments of mental health: A proposed scale. *Archives of General Psychiatry*, 1962, **7**, 407–417.

Luborsky, L., Chandler, M., Auerbach, A. H., Cohen, J., & Bachrach, H. Factors influencing the outcome of psychotherapy: A review of quantitative research. *Psychological Bulletin* 1971, **75**, 145–185.

Meltzoff, J., & Kornreich, M. *Research in psychotherapy*. New York: Atherton Press, 1970.

Morton, R. B. An experiment in brief psychotherapy. *Psychological Monographs*, **69** (1, Whole No. 386), 1955.

Nichols, R., & Beck, K. Factors in psychotherapy change. *Journal of Consulting Psychology*, 1960, **24**, 388–399.

Rogers, C. R. &, Dymond, R. F. (Eds.) *Psychotherapy and personality change*. Chicago: University of Chicago Press, 1954.

Shelley, E., & Johnson, W., Jr. Evaluating an organized counseling service for youthful offenders. *Journal of Counseling Psychology*, 1961, **8**, 351–354.

(Received October 8, 1971)

---

## Manuscripts Accepted for Publication in the
## *Psychological Bulletin*

Self-Control of Cardiac Functioning: A Promise as Yet Unfulfilled. Edward B. Blanchard (Department of Psychiatry, University of Mississippi Medical Center, 2500 North State Street, Jackson, Mississippi 39216) and Larry B. Young.

Implosive Therapy and Flooding Procedures: A Critical Review. Kenneth P. Morganstern (Department of Psychology, 314 Psychology Building, University Park, Pennsylvania 16802).

Problems in the Measurement of Cognitive Deficit. Loren J. Chapman (Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706) and Jean P. Chapman.

Behavioral Differences between Acute and Chronic Schizophrenics: Course of Psychosis, Effects of Institutionalization, or Sampling Biases? Milton E. Strauss (Department of Psychology, Wayne State University, Detroit, Michigan 48202).

Is there any Proprioceptive Feedback? Comments on Schmidt (1971). Bill Jones (Department of Psychology, University of Queensland, St. Lucia, Queensland, Australia, 4067).

Diary No.
Date 42
il. N
Bureau Ed. Research

Vol. 78, No. 6 December 1972

# Psychological Bulletin

---

## LANGUAGE ACQUISITION FROM AGE FIVE ONWARD[1]

DAVID S. PALERMO[2] AND DENNIS L. MOLFESE

*Laboratory for the Study of the Symbolic Processes, Pennsylvania State University*

A great deal of attention recently has been focused upon the language development of children between the ages of 1 and 4 years. The present paper examines the acquisition of language after 4 years of age with an eye to important developmental changes that occur after the earlier dramatic initiation of language behavior. At all levels significant advances in language were found in the literature. The evidence was interpreted as indicating that phonological, syntactic, and semantic levels of analysis are intimately interrelated and that language advances appear to be correlated with developmental periods of cognitive advance.

It has become a commonplace in recent years to point to the impressive accomplishments evidenced in the developing language of the child between the ages of 18 months, when two-word utterances begin to be heard, and 4 or 5 years of age, when the child has succeeded "in mastering the exceedingly complex structure of his native language [Slobin, 1971a, p. 1]." While the accumulated observations of the language development of the preschool child do leave one with a sense of awe, especially when viewed within the theoretical framework of recent linguistic and psycholinguistic theory (e.g., McNeill, 1970), it seems to the present authors that the emphasis upon what goes on between the ages of 18 months and $4\frac{1}{2}$ years has led to the relative neglect of development in language from age 5 onward.

Thus, without minimizing the important character of the language accomplishments up to age 5, this article attempts to show that considerable additional development occurs from age 5 to adolescence. A review of the literature indicates that the 5-year-old is far from having the equivalent of an adult native speaker's facility with the language. Scattered through the literature is evidence that at the phonological, syntactic, and the semantic levels a good deal more facility needs to be acquired before the adult level is reached.

### PHONOLOGICAL DEVELOPMENT

Perhaps the most neglected aspect of language acquisition is that of phonological development. The relative lack of research is surprising in light of the advanced state of phonological theory in general (e.g., Chomsky & Halle, 1968; Halle & Keyser, 1971) and the recently translated developmental theory of phonology advanced by Jacobson (1968).

Our review of the literature provides us with little more than anecdotal evidence about the receptive or comprehension aspects of the phonological capacities of the child. Most of the research discussed will allow us to make statements only about the productive capacities of the child. This is an unfortunate state of affairs considering the valuable insights provided by studies of comprehension at the level of syntax. While the problems associated with the study of comprehension at the phonological level may be different than those at the syntactic level, it seems useful to examine phonological development

---

[1] An earlier draft of this paper was presented at the Third Annual Michigan Conference on Applied Linguistics, January 30, 1971. Support was provided to the senior author by National Science Foundation Grant GB 8024 and to the second author by Public Service Training Grant HD 00151 from the National Institute of Child Health and Human Development.

[2] Requests for reprints should be sent to David S. Palermo, 417 Psychology Building, Pennsylvania State University, University Park, Pennsylvania 16802.

© 1972 by the American Psychological Association, Inc.

through experiments that use methodology along the lines of the Fraser, Bellugi, and Brown (1963) study, for example, to add to what we know about the productive capacities.

Beginning with the comprehensive study of Templin (1957), we have strong evidence that the articulatory skills of the child at his fifth birthday are far from complete. Although there are other studies that preceded Templin's (e.g., Poole, 1934; Wellman, Case, Mengert and Bradbury, 1931), hers is based upon a carefully selected sample, a set of 176 sounds tested in initial, medial, and final positions, and the broad age range from 3 to 8 years. The sounds were elicited by pictures, written words, or repetition of the experimenter's oral example.[3] In general, the results show a marked increase in performance from age 3 to 4, a plateau from age 4 to 5, and then significant increases in performance from age 5 to 7. At age 8 the child's articulation is essentially the same as that of the adult. Phonemes in the initial position are articulated correctly prior to those in the medial position, and the latter tend to precede correct articulation in the final position. Most of the errors at age 5 and beyond occurred on double and triple consonant blends. Only 14 sounds or sound combinations tested in initial position were failed by 25% or more of the children at this age, but some 50 other sounds or sound combinations in medial and final positions were failed by 25% or more of the sample.

It seems from Templin's data that while most individual phonemes in English are correctly articulated in most positions in words by age 5, there is considerably more difficulty in combining a series of consonants in clusters. Such consonant clusters are undoubtedly less frequent in the language, thus affording the child less opportunity to acquire the rules involved in their production. In addition, the phonological rules for production of such strings of consonants in English are surely more complex than those involving consonant-vowel clusters. It also appears from these data that the consonant features +continuant and +strident tend to be late in coming into the phonological repertoires of these children.

Recently, Menyuk (1968) reported an analysis of the role of distinctive features in children's acquisition of phonology. Her data are based upon the consonant substitutions produced by children while spontaneously generating sentences. She reported on the percentage of usage of six features for children from ages 2 years, 6 months, to 5 years. The features nasality, gravity, and voicing appear to be correctly used nearly 100% of the time by $4\frac{1}{2}$ years; the diffuseness feature comes in a bit later; and the features of continuancy in combination with stridency are not completely mastered until after the sixth birthday. Additional data collected from Japanese children in a narrower age range suggest the same rank ordering of the correct usage of these features.

A study by Snow (1964) analyzed the sound substitutions that 438 Grade 1 children, aged 6 to 8 years, made on an articulation test. While the mean number of errors for each child was small, the large sample and number of sound tests produced a substantial body of errors upon which to base generalizations. The major findings indicate that unvoiced plosives are replaced by other unvoiced plosives, and unvoiced fricatives are replaced by unvoiced fricatives. Similarly, voiced plosives and fricatives are replaced by other voiced plosives and fricatives. Semivowels are replaced by other semivowels. If a change in voicing occurs, it is nearly always an unvoiced phoneme substituted for a voiced phoneme and seldom the reverse. The most frequent errors, /f/ for /θ/ and /v/, /θ/ for /s/, /k/ for /t/, and /n/ for /m/, are the errors most frequently made in adult perceptual confusions and in substitutions made at all ages. The fact that these substitutions reflect sounds differing on but one distinctive feature is of interest, but of greater interest would be some indication of why particular single features are confused and not other single features and why single-feature confusions occur with some phonemes and not with others.

---

[3] It is of interest to note that no significant differences were obtained in the child's responses when the experimenter correctly articulated the sound in a word and the child repeated it as compared to spontaneous articulation of the same sounds. This fact is, of course, reminiscent of Ervin's (1964) report on syntactic imitation and suggests that additional study of this phenomenon, perhaps along the lines of Slobin's analysis of imitation (Slobin, 1968; Slobin & Welch, in press), may be fruitful.

The production difficulties that children evidence from age 5 to 8 seem to center on (*a*) sounds that occur in the medial or final position of a word; (*b*) consonant sounds that involve the features of continuancy and stridency; and (*c*) sounds that involve a sequence or cluster of consonants as in the case of *-lfth* (twe*lfth*), which Templin found was difficult even for 8-year-olds. As Menyuk (1971) has indicated, these production difficulties may result from the fact that the sounds involved lack clear acoustic differences that the child can distinguish. If the child fails to discriminate the acoustic cues, he will fail, in turn, to articulate them. On the other hand, if he can make the acoustic discriminations, the production errors may be attributable to difficulties with the articulatory gestures involved.

While it would be helpful to have some perceptual data to resolve this question, it would seem from what little evidence we have that the difficulty resides in the articulatory gestures and not the acoustic cues. First, it would seem from casual evidence that speech sound perception is more advanced than production and certainly begins long before production (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Second, the fact that sounds in the initial position cause little difficulty, while the same sounds in medial or final position may lead to errors, suggests that articulatory factors associated with preceding articulatory contexts may be problematical. The child may have no difficulty in producing a sound from a resting position, but when that same sound must be produced from the position of the articulators after making another sound in the sequence composing a word, the child finds the task more difficult. Even in correct articulation by adults there is a good deal more variability in articulatory movements in medial and final positions than initial position (MacNeilage, 1970). Third, the child has difficulty with consonant blends that seem to involve a rapid series of complicated articulatory gestures unrelieved by interruption of an easier vowel sound. Fourth, the errors that do occur seem to be near misses in articulation. Plosives are replaced by other plosives, fricatives by other fricatives, and so on. Children make errors that seem to result from imprecise articulation rather than incorrect articulation. Finally, the errors that are made after age 5 seem to be made on sounds such as fricatives and liquids in which partial closure in the middle of the vocal tract is required. The bilabial sounds formed with the front of the vocal tract come in early, the sounds made at the back of the vocal tract come in somewhat later, but when rather precise control of the tongue shape and position in the region from the dental ridge to the front third of the hard palate is required, production errors persist to relatively older ages. Thus, the child can identify easily, with respect to his own motor movements, the end points of the vocal tract, and he can make more easily the grosser muscular movements associated with articulating sounds at each end of the vocal tract. However, when partial closure of the vocal tract is required with constriction in various parts of the central region, more errors are made. Thus, the target that the articulators must hit seems to be smaller for fricatives and liquids, varies with context in the case of a broader range of sounds, and requires greater muscular control in the case of consonant blends.

A somewhat different approach to the understanding of phonological development was begun by Berko (1958)[4] and expanded by Anisfeld and his colleagues. Berko was interested primarily in the acquisition of the morphological rules associated with the formation of plurals, possessives, and third-person singular of the verb, and with the inflections for the past tense of regular verbs. Since all of the sounds used in producing these inflections were within the articulatory abilities of the children, the question of concern was whether the children had acquired the phonological rules that determine the particular inflectional ending for each phonological context, that is, the three allomorphs for past tense and the three allomorphs for pluralization.

Berko's procedure involved presenting the child with pictures that were referred to with nonsense or real English names and then eliciting from the child the appropriate names with an inflection. A comparison of the 4- and 5-year-old group with the 6- and 7-year-old group of children indicates that of 28 items used in her test, there were 12 items on which significantly more adultlike performance was evidenced in the productions of the older children. Pluralization of the nonsense words was generally 70% correct or better for all ages when the /*s*/ and /*z*/ allomorphs were required but less than 40% correct when the /əz/ allomorph was correct, despite the fact that 99% of the older children correctly inflected *glass* for the plural form. Similarly, when the /əz/ inflection was required in forming the third-person singular of a verb or the possessive form of a noun, performance was poor. In the case of past-tense inflection, performance ranged approximately from 50% to 85% correct when the /*t*/ or /*d*/ inflection was called for, but was less than 35% correct at all ages when the /ə*d*/ inflection was correct for the nonsense words, although the verb *melt* was correctly inflected by better than 70% of all children.

---

[4] The methodological weaknesses of this pioneering research have been pointed out by Natalicio and Natalicio (1969) and should be consulted by those interested in extensions of this work.

It would appear that in both the case of the pluralization rules and the past-tense inflection rules, the child at this age has collapsed the three rules involved into two rules, at least insofar as the productive extension of the rules is concerned. The child is able to extend the /*s*/ snd /*z*/ plural allomorphs and the /*t*/ and /*d*/ past-tense allomorphs to new forms, but he fails to use the /əz/ and /ə*d*/ allomorphs. Another interpretation is that the child has three rules in each case, but, for nouns ending in a sibilant or verbs ending in /*t*/ or /*d*/, pluralization and past tense, respectively, are formed by a zero allomorph. Alternatively, Berko favored the argument that there is only one rule in each case, a rule that /*z*/ added to a noun makes that word plural and a rule that /*d*/ added to a verb makes it past tense, and under certain conditions the /*z*/ becomes /*s*/ and the /*d*/ becomes a /*t*/ automatically by a more general phonological rule. Anisfeld's research reported below provides some additional data relative to these alternatives.

One other point with respect to Berko's study pertains to the fact that the plural of some nouns, for example, *heaf*, and the past tense of some verbs, for example, *gling*, were treated as irregular forms by a significant number of adults who offered the plural form *heaves* (42%) and the past form *glung* or *glang* (75%), which the children seldom produced. This finding suggests that children learn some irregular forms by rote and later acquire an alternative rule that is based upon a few irregular words in the language and may be extended to new words that have a similar overall phonological form to the irregular words already known. The fact that the /əz/ plural for *glass* was used correctly by 99% of the older children but not generalized to new words suggests that rote learning may be a general initial strategy in language acquisition, at least for this aspect of language, and only when the memory load or some other factor makes rote learning unmanageable does the use of more general rules come into play. Even when the rules are acquired, Berko's data on the relatively better performance with the /əz/ allomorph, both for possessive forms and for the third-person singular verbs relative to the plural form, may imply that the rules are sensitive to other factors, such as syntactic context, before they are applied only on phonological bases.

In discussing some aspects of her data, Berko suggested that children may generalize responses on the basis of phonological similarity rather than on the basis of phonetic similarity, that is, on the basis of feature similarity and not sound similarity. This point has been cleverly investigated by Anisfeld in a series of three studies that have examined some aspects of the development of phonological rules in young children.

In the first of this series, Anisfeld and Tucker (1967) revealed a number of interesting aspects of the development of the pluralization rules. An initial experiment with 6-year-old children not only established that children of this age have difficulty with the pluralization rule in which /*ɨz*/[5] is used but that they are aware both of the concept of plurality and the phonological problems involved with words ending in sounds that require the /*iz*/ plural form. These aspects of the child's conceptual as well as linguistic processes are revealed by the fact that the children tended to use numbers in their responses when they were not sure of the correct pluralization rule, usually when /*iz*/ was required. Thus, the children responded as if they knew that their pluralization might not convey the semantic concept, and therefore they preceded the word by the appropriate number to achieve the same effect.

---

[5] Anisfeld and his collaborators use the /*ɨz*/ and /*ɨd*/ phonetic symbols where Berko uses the /əz/ and /ə*d*/. Presumably this difference reflects dialect differences of the subjects used in the experiments.

A second study indicated that children seem to have a concept that the plural form of a name is the singular form with something appended. It would seem that children of this age have the concepts of one and more than one, the ability to recognize when they do not know how to correctly pluralize a word, and the concept that pluralization demands some sort of phonological addition to the word to make it plural—although they may not know what that addition should be.

Finally, the authors attempted to determine both the recognitory and productive capacities of the child at 6 years of age for the specific English pluralization rules. The results confirmed the difficulty that children at this age have with the productive use of the /*iz*/ inflection and the relative lack of difficulty with the /*s*/ and /*z*/ inflections. The mean error rate for the /*iz*/ forms was somewhat more than 50%, while it was less than 25% for the other two forms. In the recognition task, however, the error rate was high for the /*s*/ as well as the /*iz*/ form, although the overall error rate was less for recognition than production.

A study by Anisfeld, Barlow, and Frail (1968) delves specifically into the child's generalization of phonological rules in relation to the distinctive features of the sounds. Four groups of subjects, 6, 7, 8, and 19 years of age, were used. Consonant-vowel-consonant trigrams ending in /*l*/, /*r*/, or /*n*/ were presented to them as a singular name, and they were asked to choose which of a pair of plural alternatives they preferred. The plural alternatives were formed by adding to the singular form one of the following sounds: /*p*/, /*b*/, /*m*/, /*f*/, /*v*/, /*k*/, /*g*/, /*t*/, /*d*/, /*n*/, and /*ch*/.

All of the singular forms called for the /*z*/ ending so that the results were tabulated in terms of the distinctive features with the same sign that the chosen consonants had in common with the correct /*z*/ plural form. Obviously a random distribution of choices would be expected if the generalization of phonological rules played no part in affecting the choice of responses. It was found that +continuant (/*f*/ and /*v*/) and +strident (/*f*/, /*v*/, and /*ch*/) sounds were preferred significantly by the first- and second-grade subjects and the adults, although not by the 6-year-old kindergarten children. Further, it was found that the older groups tended to select the artificial plural marker with the fewest feature differences compared to the regular one. The authors argued that the features of +continuant and +strident are more important than the features +diffuse, −grave, +voice, and −nasal, which also characterize /*z*/ because it is the two features +continuant and +strident that distinguish the plural marker /*z*/ from more other consonants than any of the other features. The additional fact that voicing was not an important distinguishing feature was interpreted to indicate that pluralization rules are not formulated in terms of the voiced /*z*/ and voiceless /*s*/. Rather, a more general rule, which states that an inflectional suffix has the same sign on voicing as the preceding sound, pertains to all suffixes and therefore is not an important feature distinction in a particular instance such as pluralization. The fact that the kindergarteners did not show the same systematicity as the older subjects may be attributed to the lack of development of the phonological rules involved or, as the authors suggested, to the influence of a rule irrelevant to phonology that led them to select the second of the two alternatives presented.

A study by Anisfeld and Gordon (1968) more thoroughly examined the problem of inflectional rules. In this case, a larger number of consonantal endings were used. The subjects were first- and fourth-grade children and adults. In addition, the inflection of synthetic verbs for past tense was investigated with fifth-grade children. Otherwise, essentially the same procedures were used as in the earlier study. The use of the past-tense test was added for two reasons: First, to demonstrate that the findings of this and the previous study were not due merely to particular sound preferences regardless of the task; and, second, to show that the voicing feature is a part of a more general voicing-assimilation rule applicable to inflections in general rather than pluralization in particular.

The results showed that adults restricted their choices when possible to sounds that were +strident (/*ch*/ and /*j*/), +continuant (/*th*/) or both (/*f*/, /*v*/, and /*sh*/). The first- and fourth-grade children showed preferences only for /*sh*/, /*ch*/, and /*j*/, which make up the subclass of sounds that along with /*s*/ and /*z*/ are the functionally significant sounds in English pluralization in that they all take the /*iz*/ form for pluralization when they occur at the end of a noun. The children's rules thus seem more restrictive than those of the adults. Although they showed some tendency to prefer the /*f*/, /*v*/, and the /*th*/, found acceptable by the adults, their preferences for all sounds were less strong than those of the adults. Further evidence that the children's phonological rules are not as well developed nor as complete as those of the adults was reflected by the fact that the adults favored /*m*/ and /*n*/ endings often used in irregular forms of pluralization in English, while the children showed no such preferences, a finding that seems to corroborate Berko's results with *heaf* and *gling*.

In the case of the past-tense choices, the fifth-grade children showed significant preferences for /*ch*/ and /*j*/. The authors attributed this result to the fact that there are no other sounds relevant to past tense as there are in the case of pluralization. They argued that the /*ch*/ and /*j*/ sounds contain within them the regular past-tense markers /*t*/ and /*d*/; that is, in the articulation of /*ch*/ and /*j*/, the /*t*/ and /*d*/ sounds are part of the more complex /*ch*/ and /*j*/, respectively. In addition, the children showed a significant rejection of the /*s*/ and /*z*/ sounds that are relevant to the pluralization inflection and thus in grammatical opposition to the past-tense inflection.

In contrast to the efforts of Anisfeld to determine the phonological rules for appending a sound, in the form of an inflection, to an already complete word, Bruce (1964) attempted to examine the child's ability to analyze the component sounds comprising a word unit. His approach was to present a series of words to $5$–$7\frac{1}{2}$-year-old children (from 5 to 9 years mental age) and to ask them to delete one sound from the word and pronounce the word with the sound deleted. All of the deletions resulted in other real English words. Both the sound and position of the elision were included in the instructions for each word. The strategies employed in the task by the children at various ages are perhaps most revealing of the phonological development indicated by this task. Up to about mental age 6, the children were not able to separate sound from word. They could produce isolated sounds and words, but they were unable to take from a word one of its component sounds, regardless of whether that sound was at the beginning, middle, or end of the word. At mental age 6, the child began to be able to take a part of the word and form a new word, but the relation of the word they formed and the satisfactory solution of the task was often remote. While the 6-year-old child has begun to get the idea of sounds within words, it appears that word units merely were being substituted for word units. The last stage prior to correct performance was characterized by elision, but errors of overdeletion or incorrect deletion were prominent, and deletion from the middle of a word was particularly likely to lead to error.

These data, along with those from a study by Bryant and Anisfeld (1969) which show that children have difficulty parsing the singular form from a plural form, suggest that children take a holistic approach to words as sound units. It may be that at earlier ages the child treats whole sentences as undivided units, and as he acquires more language he parses the larger units into parts. At least in the age range from 5 to 8 years, it would appear that the word, or syllable, is the sound unit with which the child deals. He lacks the analytic abilities required to isolate and manipulate phonemic units within words. Savin and Bever (1970) presented some data that suggest that adults analyze the syllable unit prior to the phonemic unit. They argued that phonemes are abstract entities that are perceived only by an analysis of previously perceived syllables. It may be the case that young children learning inflection rules are learning new words that express semantic concepts rather than new phonemes that convert singular words to their plural form. Only at a later age, sometime after the child generalizes the */id/* and */iz/* inflectional forms, will the children make the latter analysis. The child may have developed a complex system of phonological rules by the time he is 6 or 7 years old, but he still has not abstracted the phoneme from that system of rules and mastered the complex relationships which it has to acoustic stimuli and to articulatory movements.

Another bit of data that may relate to the same point is provided by Warren and Warren (1966). Their results indicate that compared to college-age subjects, 5- and 6-year-old children do not hear as many verbal transitions nor as many different verbal forms in a verbal transformation task. When children do report transformations, they tend to involve several sound changes at once, and sometimes additional sounds, rather than gradual single sound changes more characteristic of the older subjects. Children from 8 to 10 years of age tend to respond in a manner more similar to the adults.

Warren and Warren (1966) interpreted their results to indicate that the younger children organize the repeating stimulus in terms of individual phonemes because they report sound sequences not permitted in English. Just the opposite interpretation might be indicated since the adults tend to change only single phonemes when they report transformations while the children report fewer but larger transformations. Children cannot break up the word unit. When they do report transformations, the form of the word changes by several phonemes. These results do not seem contradictory to the hypothesis that words or syllables are the perceptual units for young children—as Savin and Bever have argued that they are for adults—and that, in contrast to the adult, the child prior to about 8 years of age has not yet developed the conceptual ability to abstract the phonemes from the word unit.

In summary, it seems that there are ample data here to allow us to agree fully with Chomsky's comment that "It is by no means obvious that a child of six has mastered his phonological system in full—he may not yet have been presented with all of the evidence that determines the general structure of the English sound pattern [Chomsky, 1964, p. 7]." It seems equally clear that the area of phonological development deserves a good deal more attention both before and after age 5 than it has received from researchers to date. In particular, the examination of the perceptual as well as the productive aspects of the system requires further research effort. Perhaps an expansion and elaboration of the sound-discrimination task, reported only briefly by Templin (1957), would be a valuable place to start. In addition, the relationship between acoustic characteristics of sounds and the production difficulties of the child might be revealing in light of Liberman's (1966, p. 193) comments indicating that */s/*, */sh/*, */l/*, and */r/* are some of the most stable acoustic phonemes, but they tend to cause articulatory problems for the child. Finally, little or no research has been done on the prosodic aspects of the phonology of children this age.

## SYNTACTIC DEVELOPMENT

Turning now to syntax, we find considerably more research available and a large number of studies conceived within the general transformational-generative framework. The results of these studies have frequently been interpreted to support the contention that "All the basic structures used by adults to generate their sentences can be found in the grammar of nursery school children [Menyuk, 1963, p. 419]." The interpretation turns, of course, upon what is meant by the phrase "basic structures." Without trying to quibble about what is or is not a basic structure, some data suggest that important syntactic advances occur long after the child has passed his fifth birthday.

The first set of studies that is examined here attempted to evaluate syntactic development by the procedure of collecting a corpus of language from children of various ages, followed by an analysis of that corpus in terms of a set of grammatical rules that could be used to describe it. The rules that describe the corpus are, in turn, compared to the rules presumed to account for adult language, and evaluations of development are made. The methodological and interpretative difficulties with such an approach are, of course, serious, although they are not discussed here.

The studies of Menyuk (1963, 1964, 1968), Loban (1963, 1966), and O'Donnell, Griffin, and Norris (1967) are particularly extensive examples of this type of approach. Menyuk examined the language of children from 2 to 7 years of age. Approximately 80 to 120 sentences were collected from each child, although no indication was given of how the sentences were selected from the running speech of the child. On the basis of grammars written to describe the sentences of the children, Menyuk concluded that nursery-school children have completed the phrase structure and morphological levels of grammar. Her analyses focused upon transformation rules in which some developmental trends were observed. Menyuk suggested that nearly all transformations used in adult language are present in some of the nursery-school-aged children, but even the first graders, as well as the kindergarten children, in the age range of 5 to 7 years failed to exhibit full development of the auxiliary *have*, participial complement, iteration, nominalization, pronominalization, and conjunctions with *if* and *so*. In addition, there were some 17 types of restricted transformations used only by children and not apparent in the language of adults. For example, noun phrase redundancy, as in "She took it away the hat," actually increased significantly in frequency from nursery school to Grade 1. Finally, there were some structures used by adults that never appeared in the children's language at any age level although these were not specified. Little can be said about whether the latter failed to appear because of lack of opportunity or lack of the competence to produce such structures.

Loban (1963, 1966) conducted a longitudinal study of language development over a 10-year period for a group of 220 children carefully selected in terms of socioeconomic level and a number of other variables. The study began when the children were in kindergarten and continued through ninth grade. Each subject was individually interviewed annually, and his spoken responses recorded. The results indicate that as children get older their speech performance improves, as indicated by decreases in incomplete syntactic structures, increases in the variety of structural patterns used, and greater variation in the structures within sentences in terms of vocabulary, positions of phrases (such as adverbial modifiers), nominalizations, and so on.

In a similar study, O'Donnell et al. (1967) collected samples of oral and written language from 5- to 14-year-old children. The analyses of the data were similar to those of Loban but based upon terminal syntactic units (T units) used by Hunt (1965) in an earlier study of children's written language. Simple or complex sentences were defined as T units, but a compound sentence was analyzed in the smaller T units of which it was composed.

The length of the T units increased from about 7 words for kindergarteners to about 10 for the seventh graders. Of more interest, however, is the evidence for two periods in which sudden changes in performance appear to occur. Between kindergarten and first grade and between fifth grade and seventh grade are developmental periods when large increases in new grammatical constructions, or sudden increases in the use of constructions previously evidenced at low frequencies, and high error rates on some kinds of constructions seem to occur. Mean number of sentence-embedding transformations within T units increased significantly at both of these transition periods. Marked increases in nominal, adverbial, and coordinate constructions also occurred at both of these developmental points. Nominals containing adjectives and prepositional phrases particularly increased from Grade 5 to 7. Finally, frequency of coordinate nominals and coordinate predicates increased significantly from Grade 5 to 7.

In addition, O'Donnell et al. reported that only 2.6% of the seventh graders' utterances (excluding garbles) could not be accounted for by 1 of 11 common structural descriptions (e.g., subject-verb-object, etc.), while 12.5% of the kindergarten children's utterances fell outside of these categories. Most of the excluded utterances, however, did not involve gross grammatical errors. The subject-verb and subject-verb-object structural forms accounted for approximately 77% of the grammatical utterances of kindergarten children and nearly 90% of the seventh-grade children's utterances. Passive constructions accounted for less than 1% of the sentences of any of the grades, although there was a slight rise in these forms with age. These results are consistent with those of Loban and fit well with the data reported by Goldman-Eisler and Cohen (1970), who found that only 4% of adult utterances were passive, while 87% were of the simple, active, affirmative, declarative form.

In summary, the Menyuk (1963, 1964), Loban (1963, 1966), and O'Donnell et al. (1967) research provides a general overview of further language development in the child after 5 years of age. The overall results suggest that there is a general but gradual consolidation of language structures from kindergarten to seventh grade but also abrupt shifts in performance, which occur between kindergarten and first grade and between the fifth and seventh grades. More research focused upon these two periods may be of particular interest in revealing what is happening at these ages. It may be that the child is acquiring rules for different syntactic structures at these ages and that these rules affect and disrupt other structures that the child previously has dealt with in a competent manner. Language is an integrated system in which a change in one structure cannot help but affect other structures within the system. Examples of this latter phenomenon are seen in the data reported by Chomsky (1969) and Kessel (1970) using more controlled experimental procedures in which specific linguistic constructions were investigated.

Chomsky (1969) studied aspects of comprehension as a function of a number of factors including sentence complexity, selectional restrictions, and subcategorization restrictions. She was concerned with the developmental control over deep structure–surface structure relations when particular lexical items do not conform to rules that frequently relate surface structure to deep structure. She focused her attention upon the lexical entries *ask*, *tell*, and *promise*; pronominalization; and the syntactic comprehension of sentences in which the surface structure subject is not the deep structure subject. Her subjects ranged in age from 5 to 10 years. A Piagetian interviewing method with structured tasks and specified initial questions was used with more open questioning to evaluate the child's comprehension of the language-task relations.

In the case of *promise*, the so-called minimal distance principle applicable to most verbs is violated. In sentences that conform to the minimal distance principle, the noun phrase that immediately precedes an infinitive complement verb is the subject of that verb. Thus, in "John wanted Bill to leave" it is Bill who does the leaving, but in "John promised Bill to leave" it is John who does the leaving.

Chomsky's results indicate that while even the youngest children in this age range seem to know the meaning of *promise*, although there is some doubt as to the adequacy of Chomsky's check on this, they show systematic developmental trends suggesting that they do not correctly comprehend sentences involving both *promise* and the minimal distance principle until they are about 8 years old. The children appeared to pass through a stage in which they first applied the minimal distance principle to all cases across the board. Next, as children began to recognize exceptions to the rule, they made errors with sentences that followed the minimal distance principle as well as those that did not. In the third stage, the children finally straightened out the cases to which the minimal distance principle applied but continued to mix their responses to the exceptions so that sometimes they treated *promise* appropriately and sometimes overgeneralized the principle. Only at the fourth stage was errorless performance achieved in all cases. While there was a good deal of variability with respect to the ages of the children who achieved each stage, the sequence of achieving the stages was orderly and nontransitive. Similar results were found when *ask* and *tell* were used in another experimental task in which *ask* was used both as a command and a question. *Tell* consistently follows the minimal distance principle, while *ask* is inconsistent and does not when used in the command or request form. Development of the correct interpretation of *ask* in the command form was correlated with the development of *promise*, and similar stages of confusion with *tell* were apparent. Comprehension of *promise*, which consistently violates the minimal distance principle, always preceded the comprehension of *ask*, which is inconsistent with respect to this rule.

Kessel (1970) also investigated the *ask–tell* relationships with a group of children ranging from kindergarten to fifth grade. His results were highly correlated with Chomsky's (1969). The major difference was that Kessel's subjects tended to achieve the various stages described by Chomsky at somewhat earlier ages. Like Chomsky, Kessel found that children passed through an invariant and nontransitive sequence in their acquisition of *ask* and *tell*.

Chomsky's (1969) examination of the comprehension of sentences in which the surface structure subject is not the deep structure subject made use of the child's response to the sentence "Is this doll easy to see or hard to see?" in the context of a blindfolded doll. Her results indicated that 5-year-old children tend to interpret the sentence to mean that the doll is hard to see, that is, that *doll* is the deep structure subject of the sentence rather than the object of the infinitive. By 9 years of age no child incorrectly interpreted the question.

Both Kessel (1970) and Cromer (1970) have challenged Chomsky's (1969) methodology in this task. Kessel argued that the blindfold is an unnecessarily distracting cue in the task and that poor performance may be attributed to the fact that the younger children did not recognize that the blindfold was irrelevant to the question. Further, Kessel noted that Chomsky did not test comprehension of superficially similar sentences in which the deep and surface structures are the same.

Kessel used a simulated hide-and-seek game in which the child responded to eight declarative sentences. Half of the sentences were of the form "Lucy was sure to see," and half were of the form "Lucy was impossible to see." The results indicated that 6-year-old children have little difficulty with sentences in which the deep and surface structure subjects are the same, but they do have difficulty in the comprehension of sentences in which the deep and surface structure subjects are not congruent. The children's errors indicated that the difficulty in sentences of the latter form is manifest in the assignment of the incorrect subject to the infinitive verb. In line with Chomsky's results, Kessel found that by 9 years of age children have little difficulty with such sentences. Again, a nontransitive sequence of stages was found in the acquisition of sentences of this type.

Cromer (1970) examined the comprehension of sentences in which the nouns and verbs were held constant, but adjectives, such as *eager* and *easy*, determined the relation of the deep and surface structure subjects. His subjects used hand puppets to act out the sentences uttered by the experimenter. Cromer used four adjectives that unambiguously require the noun to be interpreted as the subject of the sentence, four adjectives that unambiguously require the noun to be interpreted as the object of the infinitive verb, and four adjectives that render the sentences ambiguous. The results indicate that children of mental age up to 5 years, 7 months, tend to interpret all sentences as if the noun were the subject of the sentence; between 5 years, 9 months, and 6 years, 6 months, the children are in a transition stage in which the first type of sentence is correctly interpreted, but there are both correct and incorrect interpretations of the second type of sentence; and children over mental age 6 years, 8 months, interpret both the first and second sentence types correctly and interpret some of the ambiguous sentences one way and some the other. Cromer divided his groups by mental age rather than chronological age because the apparent developmental trends seem to emerge more clearly. This may be an important finding that others should note more carefully in future studies plotting the course of language acquisition.

Cromer also found that all but two of the youngest children had no difficulty with comprehension of passives, indicating that these children knew that what appears to be the surface subject in sentences need not necessarily be the deep structure subject. Finally, he presented the children two nonsense words in sentence frames, which required them to be differentiated with respect to what was to be taken as the deep structure subject and then tested for comprehension of the two nonsense words. Children in the first stage failed to make the differentiation; some of the children in the transition stage correctly interpreted the two new words, but most of the children did not; and, finally, those children who were in the third stage correctly interpreted the two new nonsense words on the basis of hearing them used in only one repeated sentence. Thus, it would appear that Chomsky's results have some generality when tested under different conditions.

In the case of pronominalization, Chomsky found that children learned the nonidentity restriction for pronominalization between 5 and 6 years of age, while seven-year-olds were able to handle, in addition, the identity case. Thus, kindergarten children were able to comprehend sentences such as, "He knew that John was going to win the race," in which *he* and *John* refer to two different persons. Not until the children were older, however, did the identity situation, as in "John knew that he was going to win the race"—where *John* and *he* refer to the same person—no longer present any difficulties. Chomsky (1969) argued that since the second form requires a more complex set of rules, because of the ambiguity involved, it would be acquired later. Since the ability to deal with pronominals in these situations is basically dependent upon two rules, she theorized that this would account for such structures being acquired relatively early, while other constructions such as the *ask–tell* distinction would require more time to be mastered because a more specific and complex set of rules is necessary to account for them.

Other investigators, however, have found that difficulties with pronominal forms persist into the junior-high-school level.[6] Loban (1963), for example, found a marked increase in errors of pronominalization at the seventh grade which did not decrease to the performance level of the sixth graders until after grade 9. Chai (1967), in a controlled experimental procedure, also reported that difficulty in comprehending pronominal referents in compound sentences extends into the junior-high-school range. It seems that the development of structures involving pronominalization are far from complete by 5 years of age, as evidenced in both production and comprehension capacities of children well beyond that age.

There are a number of studies that have focused on comprehension and production of question, negative, and, particularly, passive sentences by children in the age range we have been considering. Slobin (1966) reported a rather extensive developmental study of the comprehension of various sentence structures including passives. His data for 6-, 8-, 10-, and 12-year-old children and for adults indicate a developmental trend toward increasing performance in comprehending passive, negative, and passive-negative sentences. Unfortunately he reported only latency data, and the more revealing error data collected are not reported in the published paper. It is of interest to note, however, that performance on passive sentences is superior if the subject and object of the passive sentence are not reversible, a result that was especially clear for the 6- and 8-year-old children. In fact, the nonreversible passive sentences were nearly as easy as the simple, active, affirmative, declarative sentences. Slobin argued that the difficulty in keeping track of which noun is the actor is eliminated on a semantic basis in the nonreversible passive sentence.

Turner and Rommetveit (1967a) provided support for Slobin's hypothesis. In a developmental study of active and passive sentence imitation, comprehension, and production, they found that 94.5% of the errors made in the various tasks involved inversion of the two noun elements. Most of the errors occurred on the reversible sentences, but a surprisingly large percentage (32.6%) occurred on the nonreversible sentences, which suggests that even with semantic support, children in the age range from 4 to 9 years continue to have difficulty with comprehension and production of passive sentences. Not until second grade (mean age 8.11 years) was performance with reversible passive sentences better than 60% correct on either comprehension or performance tests, although the kindergarten children (mean age 5.87 years) performed at approximately the same level on nonreversible sentences.

---

[6] L. S. Golub, W. C. Fredrick, & S. L. Johnson. Development and refinement of measures of linguistic abilities. (Working Paper No. 33) Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, 1970. Golub et al. used a multiple-choice format to test comprehension of sentences containing a pronoun and found that a large percentage of fourth and sixth graders had difficulty identifying the referents of pronouns in sentences, particularly when relative pronouns were involved.

A second study by Turner and Rommetveit (1967b), using the same subjects as in the previous study, tried to provide contexts in which children would be likely to use the passive structure. The children were presented pictures and encouraged by the experimenter to describe the pictures with passive sentences by focusing attention upon the acted-upon noun rather than the actor noun.

The results indicated that spontaneous use of passive forms was minimal. Not until first grade (mean age 7.00 years) was it possible to induce the children to give more than 50% passive sentences even after an example had been given and the acted-upon object was shown first in the picture. When questions were asked about the object, kindergarteners gave approximately 50% passive responses. Those who did use the passive forms tended to show advanced linguistic abilities on the independent measure of imitation, comprehension, and production reported in the previous study. Thus, there appear to be individual difference variables involved, and most children do not seem to have command of passive forms until after age 5. There was a tendency to give passive constructions more frequently when nonreversible actor-object relations were involved.

Hayhurst (1967) also found that children between the ages of 5 and 9 have a great deal of difficulty producing passive constructions even when the experimenter provides passive examples in the context of describing a picture. If the passive sentences are negated, the child's difficulty is compounded, but if the passive is truncated, that is, without an expressed actor, the child's difficulty is reduced. Under no condition, however, was performance as high as 50% correct for the 5- and 6-year-old children in Hayhurst's study.

Another study confirms the difficulty that children have with passive constructions (Gaer, 1969). Not until 6 years of age did Gaer's subjects show comprehension of passives that was well above chance, and Gaer's production test yielded little evidence of ability to produce passives at that age. Gaer also found that 5-year-old children have considerable difficulty in comprehending and producing sentences with embeddings, particularly when there are two embeddings or a single center embedding.

In summary, it would appear from these data that the 5-year-old child is just beginning to fully comprehend the passive construction and very seldom uses the passive in his own spontaneous speech. The various experimental attempts to elicit passive constructions yield results that make it difficult as yet to judge when production capacity is well developed. Clearly, there are marked individual differences that are, perhaps, more evident with this construction than most others that have received research attention. It may be fruitful to examine variables that are considered to affect language acquisition in general, with the acquisition of passive forms. The technique, used with adults by Wright (1969), of giving sentences in passive (or active) form and then asking questions about the sentence in active (or passive) form might be used effectively in revealing some of the difficulties that children have with passive sentences.

Finally, in this section we consider some data pertaining to connectives, which could, perhaps, be discussed in the next section on semantics since there seems to be a strong semantic component involved in the results, although the specific words of interest are usually considered function rather than content words. A study by Katz and Brent (1968) provided some information about the comprehension and production of the connectives *because*, *then*, *therefore*, *but*, *although*, and *and*. Their data are based upon a corpus of spontaneous speech, responses to pairs of sentences in which various connectives were contrasted, and explanations for choices made in the latter situation. The subjects were first- and sixth-grade children and a college student group. Some of the data clearly suggest that the meaning of *because*, *then*, and *therefore* changes between first and sixth grade. While the first grader may use these words in his spontaneous speech, it would appear that the temporal relations of *because*, for example, are better understood than the causal ones, and the younger children probably do not have more than a sequential, as opposed to a causal, meaning for *because*. In other words, all three of these words are used as if they were marked semantically as *then*, with no causal relations implied. In addition, it appeared that when the adversative connectives *but* and *although* were used, children in the first grade showed little evidence of comprehending such constructions, and the sixth graders, while better in the identification of sentences correctly using the words, showed little ability to account for their choice. Finally, there was a developmental trend revealing an increase from Grade 1 to Grade 6 in the preference for the linguistic order of clauses to mirror the temporal order of cause and effect events. Thus, sixth graders seemed more aware than first graders of the cause and effect relations in such sentences as, "The benches were wet, we did not sit down" and showed a marked preference for stating the sentence in that order rather than in the order, "We did not sit down, the benches were wet." The college students showed no preference for order. These results may reflect a general cognitive developmental awareness of cause and effect by the older children and, at a somewhat more abstract level, may reflect the same kind of perceptual linguistic interrelationship showed by Huttenlocher in her studies (Huttenlocher, Eisenberg, & Strauss, 1968; Huttenlocher & Strauss, 1968) and Bever in some of his work (Bever, 1970).

In a series of studies concerned with logical thinking within a Piagetian framework, Neimark reported some surprising results related to the connectives *and* and *or*. In the first study (Neimark & Slotnick, 1970), children in Grades 3 through 9 and college students were presented with a sequence of 16 problems dealing with class inclusion and exclusion, class intersect, and class union. The results indicate that none of the subjects had any difficulty with class inclusion and exclusion, only the third-grade children had any difficulty with class intersection, but only the college students achieved success on the majority of the union problems. Children in Grade 9 responded correctly only 30% of the time when given instructions to select A or B, and performance was still poorer when the presumably clearer instruction to select A or B or both was given. According to the authors, these results are not culture bound since they replicate findings with Japanese children. An analysis of the errors revealed that most of the children interpreted *or* as *and*, a result also apparent in a study by Shine and Walsh (1971).

A second study (Neimark, 1970), conducted with children in Grades 9 through 12, indicated that correct performance on union problems increases steadily from Grade 9 to 12. A particularly interesting finding comes from the children's descriptions of the union sets in one form of the test used. It would appear from the children's own efforts to describe the sets that they do not use the word *or* spontaneously, at least under these circumstances. Only 6 of 39 subjects used *or* in describing union sets. Instead of using *or* they used *and* in describing the sets. They did not even use *except* or *all but* with any frequency. This failure of children as old as high school age to be able to comprehend a word as superficially simple and frequently used as *or* suggests that syntactic and/or semantic development extends over an unexpectedly long period and that it must be carefully examined throughout childhood.

Along similar lines, Olds (1968) has developed a game for children which cleverly allows exploration of the comprehension of various syntactic forms. He has shown, among other things, that children as old as 9 years do not comprehend the conditional *unless*. Children of age 9 and younger interpret *unless* as if the word *if* had been used.

Finally, a study by Goodglass, Gleason, and Hyde (1970) may be mentioned here. As a part of their study, they tested normal children between the ages of 3 and 10 years for comprehension of prepositions. In one test the subject was shown sets of three pictures and asked to identify, for example, the picture with the girl standing *behind* (as opposed to *in front of* or *next to*) the car. In the other test, the subjects were asked to indicate which of two sentences involving a contrast in prepositions was a better description of a picture.

Performance was relatively good on the first test and showed a steady increase from about 73% correct for the 3- and 4-year-old children to about 97% for the 10-year-old children. The biggest increase occurred between 5 and 6 years. On the second test, however, performance was near chance level until the age of 6 when performance jumped to a level indicating full comprehension of the correct prepositional forms. Further investigation of the comprehension of prepositions will be required before firm conclusions may be drawn.

The results of these studies of syntactic development point to a close interrelationship between general cognitive development and the comprehension and production of syntactic forms. Slobin (1971b) recently has suggested that this is the case, and he has outlined a model beginning with semantic intentions, which are based upon the general cognitive developmental level of the child, and relating those semantic intentions to the means for expressing them in a linguistic form. Particular linguistic forms are not comprehended nor produced until the underlying cognitive aspects are developed. Once such cognitive development has occurred, the child will look to the language for the means to express the new cognitive structures. Thus, examination of the language development of children, which reveals errors of various kinds at different ages, indicates either that cognitive development has not reached the point where the linguistic forms in question have any meaning for the child or the child has not as yet discovered the appropriate linguistic means for expressing the meaning he does know, a point that McNeill (1970) discussed in terms of weak and strong linguistic universals. In some cases it may be quite easy to distinguish which problem the child is having, while in others a theory of cognitive development may be a more important prerequisite to the understanding of language acquisition. For example, in the previous section the research of Anisfeld and Tucker (1967) clearly indicated that 5-year-old children had the cognitive concept of plurality but did not always know how to express it in a linguistically correct form. Slobin's examples with bilingual children make this point in a particularly convincing fashion. On the other hand, the review of the studies of pronominalization present an unusually complex picture in which the child seems to grasp pronominal constructions at one age only to lose them and reacquire them again, with final errorless performance coming relatively late in language acquisition. Surely the child acquires the concept of the pronoun as a noun substitute rather early. The erroneous linguistic structures that occur at later stages result either from attempts to express more complex cognitive relations in which the pronoun and its noun substitute become confused in the form of linguistic expression, or from the development of new cognitive structures to which pronominalization is differently related.

While the research reviewed in this section reflects a steady development of linguistic form from age 5 to adolescence, there is an indication across a number of the studies that the periods between 5 and 8 years and between 10 and 13 years are marked by instability in linguistic development, and these periods are followed by growth to new levels and subsequent stable linguistic performance. It may not be coincidental that these are precisely the periods in cognitive development marked by Piaget (1970) as transition points from preoperational thought to concrete operations in the first case and from concrete operations to formal operations in the second case.[7] It is during these two periods that large increases in new grammatical constructions and high error rates on some kinds of constructions are reported. The minimal distance principle is worked out during the first of these periods, as is the passive construction. Pronominalization performance advances during both of these periods. The work of Katz and Brent (1968) provided some specific hints about the relation between cognitive development and linguistic expression. The indication of a shift in the use of *because* from a mere expression of time sequencing to cause and effect relations, and the ordering of phrases in sentences, are two examples that seem to indicate that even when the same surface structure forms are used, the underlying syntactic and semantic relations may change as the child develops cognitively. Finally, the research of Neimark and her associates on union problems appears to fit very nicely with Piaget's notions of formal operations and gives additional support to the cognitive basis of language development.

---

[7] Sinclair-de Zwart (1969) has argued along similar lines and has noted the fact that the beginning of structured language occurs as the child's cognitive development shifts from the sensorimotor to the preoperational stage at about 18 months to 2 years of age.

## SEMANTIC DEVELOPMENT

Since the research that might be considered relevant to semantic development has not been integrated by any particular theoretical point of view, it is difficult to interrelate it or to determine which kinds of data are of importance in the evaluation of the child's semantic system. To begin with, it is obvious that additions to the vocabulary are made throughout the life span. The fact that we add a new word to our lexicon, however, may have relatively little effect upon our semantic system. It is probably insignificant to a consideration of the semantic system if a new word merely acts as a label or shortcut for expressing something that it is possible to express using other words. For example, an unmarried male expresses adequately what is meant by bachelor, and the addition of the latter word to the vocabulary says little about language development. Thus, developmental studies of vocabulary growth, while gross indexes of semantic development, may not be very revealing of the system itself.

The problem, of course, is to identify the significant aspects of semantic change as the child develops. This problem is not limited to the nature of the semantic features in the narrow sense but must also include the interword semantic restrictions. We must examine not only the acquisition of the semantic features that mark the words *wish* and *want*, for example, as nearly synonomous but also the semantic restrictions and syntactic restrictions that allow us to say, "The man wishes that she would go home," but not, "The chair wishes that she would go home" or "The man wants that she would go home." Thus, the expansion of the semantic system involves not merely the meaning of isolated words, however described, but the interrelations of those words with others. Research in this area, therefore, needs to be devoted to a consideration, in studies of both comprehension and production, of the development of semantic markers and selection restrictions, both syntactic and semantic.

Turning now to the data, we devote our attention to some evidence for the evolving meaning of lexical items in the dictionary, avoiding research reviewed in other discussions of the topic (e.g., McNeill, 1970). A number of studies scattered through the literature indicate that the elementary-school-aged child may have a substantial set of words in his vocabulary that have different meanings than those same words have for adults. Asch and Nerlove (1960), for example, investigated the development of what they call "double function" words. Such words as *bright*, *hard*, and *sweet*, for example, have both a physical and a psychological meaning in the sense that they may refer to both physical characteristics of objects and to psychological characteristics of persons. The authors studied the development of these two functions of such words in children between the ages of 3 and 12 years. In a Piagetian type of interview situation, the children were presented a series of objects and asked if any of the objects could be called *sweet*, or *cold*, or *crooked*, etc. If the child could identify correctly the object associated with the words, he then was asked if a person could be "sweet." If the child used the word in the psychological sense, he was asked about the relation of that meaning to the physical meaning.

The results indicated that the 3- and 4-year-old children could correctly use the words with respect to the physical objects, but, with the possible exception of *sweet*, they had no idea that the words could be applied to people, and, in fact, they denied specifically that these words could be so used. The 5- and 6-year-old children showed little evidence of change from the performance of the youngest children. In the 7- and 8-year-old group, however, about half of the words were correctly applied to people, but the children could not relate the physical meaning of the word to the psychological meaning with any great success. By 9 to 10 years of age, half of the children could correctly use all of the words in both contexts, and the ability to indicate the relation between the two uses began to be evident. While all of the 11- and 12-year-old groups did not perform perfectly, they did show marked advances in the comprehension of the dual function of the words.

The authors concluded that children first master the object reference of double function words, then the psychological sense of these terms as independent meanings, and, finally, the dual or relational aspects of the words are acquired. It appears that the two related meanings of these words are acquired independently by the children first, and only later are the common features of the meanings integrated. If the semantic markers associated with the two meanings of each of these words are functionally independent and, therefore, entered separately into the dictionary at first and only later unified, then these results seem most consonant with what McNeill (1970) has called vertical development of the dictionary. The latter hypothesis is that lexical items enter the dictionary list with a nearly complete set of semantic markers, and semantic development consists primarily of establishing the relations among entries through the vertical organization of common markers.

The results also may be interpreted as an indication of the relation between semantic development and general conceptual development of children who give other indications of moving from the ability to conceptualize primarily in terms of concrete operations to more abstract levels at about 11 or 12 years of age (Piaget, 1970). Additional evidence for the development of more abstract conceptualization of words, which may be conceived as the vertical organization of semantic markers, may be found in some developmental studies of word definitions (e.g., Al-Issa, 1969; Wolman & Baker 1965), which indicate that concrete and functional definitions of words give way to abstract definitions at about 10 to 12 years of age. Similarly, in some recently reported research, Anglin (1970) emphasized the concrete to abstract developmental trend in semantic development.

At about the same time that Asch and Nerlove were studying the developmental convergence of the meanings of words, Ervin and Foster (1960) were examining the developmental differentiation within sets of words that are frequently correlated in usage but that are not synonymous. Children in Grades 1 and 6 were presented with objects that could be described in terms of weight, strength, and size and pictures of faces that could be described as *happy*, *pretty*, *clean*, and *good*. The materials within each set were presented in pairs such that one of the dimensions was contrasted in the pairs and the other dimensions held constant. The children were asked about all of the dimensions. The results revealed that both first and sixth graders indicated that the dimensions held constant also varied in addition to the one actually contrasted in the objects presented. The effects were so strong with the words *good*, *pretty*, and *happy* that the authors concluded that these three words are interchangeable synonyms for first-grade children. The tendency to treat the *big*, *strong*, *heavy* set as synonyms was less strong, and in both cases the effects decreased with age. In contrast with the results of the Asch and Nerlove study, these results point to the interpretation of what McNeill calls the horizontal development of semantic markers rather than vertical development. These sets of common words appear to have entered the dictionary without all of the semantic markers required to differentiate among them. As semantic development progresses, additional markers are associated with each entry so that eventually words that were once treated as synonyms are differentiated.

A more recent study by Lumsden and Poteat (1968) points to the same interpretation. They were concerned only with the concept of "bigger" in children 5 and 6 years of age and a high school control group. They presented children with pairs of geometric figures in which the area was equal but the vertical dimension was varied, and the child was asked which of the two was bigger. For example, one pair consisted of an 8 × 8 inch square and a 13 × 4.8 inch rectangle with a vertical axis of 13 inches. Two additional pairs were used in which the area as well as the vertical dimension varied.

The results indicated that the 5- and 6-year-old children selected the object with the greatest vertical dimension in approximately 72%–97% of the pairs. Even when the area of the alternate object was four times as great as the object with the greater vertical dimension, 85% of the children selected the smaller object with the greater vertical dimension. Since rectangles were used and therefore the vertical lines did not bisect the horizontal lines, there is little likelihood that these results are attributable to the vertical–horizontal illusion. This illusion actually increases from 6 years of age to adulthood when bisection is not involved (Fraisse & Vautrey, 1956).[8] If we assume that these results are attributable to a lack of the appropriate semantic markers on the lexical entry *bigger*, rather than to some factor associated with the perceptual salience of the vertical dimension in the experimental task, then these data may be interpreted as supporting the hypothesis of horizontal semantic development.

Finally, another study that attacks the same problem with still another technique is reported by Bradshaw and Anderson (1968). Their interest was in adverbial modifiers. They investigated the problem by asking their subjects to compare each of nine adverbial modifiers with all of the other modifiers using a paired-comparison procedure. The modifiers were *slightly*, *somewhat*, *rather*, *pretty*, *quite*, *decidedly*, *unusually*, *very*, and *extremely*, and all were modifying the word *large*. The subjects were drawn from Grades 1 through 6, 8, 10, 12, and college. The procedure proved to be highly reliable at all grades, and developmental differences in meaning were evidenced at the younger grade levels. Specifically, the modifiers *slightly* and *somewhat* were neutral modifiers for the children in the first two grades. *Slightly* shifted to the most extreme position by Grade 4, and *somewhat* moved into its adult position next to *slightly* more gradually and was not located where adults place it until Grade 8. *Very*, at the other end of the continuum, was considered the equivalent of *extremely* by first graders and gradually shifted toward the neutral position in ranking and was located at the adult position by fifth-grade children. The fact that *slightly large*, for example, is considered by the first-grade children as meaning larger than *rather large*, *large*, or *quite large* implies that the markers associated with the adverbial modifier *slightly* are either different or incomplete, relative to those that the adult has. While these data seem to support the horizontal semantic development hypothesis of McNeill, they also suggest the possibility that a third hypothesis may be tenable. It is conceivable that the young child may enter incorrect markers for some lexical entries in the dictionary, and semantic development may consist of acquiring enough knowledge about entries to not only add markers as in horizontal development but in some cases subtract markers associated with particular entries.

In summary, our examination of the literature in this area has been far from complete, but the data reviewed seem to leave us in the same position that McNeill found himself in evaluating the characteristics of a set of developmental data dealing primarily with word associations, word recognition, and recall of word strings. Both the horizontal and vertical semantic developmental hypothesis seem to be supported by parts of the data. Perhaps both processes take place, but some additional research, focused on attempting to directly test the hypothesis, is needed since it is always difficult to evaluate research within one framework when it was collected within another.

## CONCLUSIONS

In conclusion, while this review has not been exhaustive, there are a number of general impressions that emerge. First, there are rather clear indications that language development is far from complete when the child reaches his fifth birthday. Regardless of whether we examine phonology, syntax, or semantics, we find data already in the literature that indicate that further development is required before adult language competence may be reached. Furthermore, there are numerous hints in the literature reviewed that additional research will reveal other areas, not yet fully explored, in which development is incomplete. For example, at the phonological level there are indications that adults have a set of rules that subsume subsets of the so-called irregular inflectional forms. Young children, on the other hand, evidence no such rules and seem to acquire the irregular forms by rote memory. Presumably after enough experience with enough examples within each subset of irregular forms, the child infers rules that govern subsequent inflections for new words. It also may be possible that a hierarchy of rules is acquired such that in this case, for example, the regular or more frequently used rule is applied first, but other rules are available if the regular rule is inappropriate.

---

[8] We are grateful to Charles N. Cofer (personal communication, May 1971) who pointed out that this illusion might be relevant here.

Second, this review suggests that in the analysis of children's performance it is becoming more difficult to clearly distinguish phonological, syntactic, and semantic aspects of language as separate levels that can be studied independently; for example, children seem to be aware of the semantic aspects of inflections, and when they do not know the phonological rules required to convey a semantic concept, they use some other method of conveying the concept. At the syntactic level the so-called function words, which in and of themselves are semantically empty, do involve meaning, and the analysis of the comprehension and use of these words cannot treat them from a syntactic point of view alone.

Third, the interrelationship of phonology, syntax, and semantics with overall cognitive development was continually apparent throughout our consideration of the research reviewed. It is clear that a theory of language development must be embedded within the larger context of a theory of cognitive development. While cognitive development may precede its expression in language, it is obvious that the reverse is unlikely. Thus, for example, the use of plural inflections must be preceded by a concept of number, and the full comprehension of the logical connective *or* must be preceded by the concept of set union. In addition, the suggestions in the data that the age ranges of from 5 to 7 and from 12 to 14 may be important transition points in language development are also the age ranges in which other more general cognitive changes are assumed to occur, at least within a Piagetian framework. White (1965) has documented many other changes that occur in various cognitive skills at the 5–7-year age range. Further exploration of the 12–14-year age range may prove fruitful both from a cognitive and a language point of view. The general point that cognitive development and language development are related intimately has also been emphasized strongly by Slobin (1971b) and Olson (1970) as well as Piaget (Sinclair-de Zwart, 1969). In this connection, mental age has been shown to be related closely to performance measures of language development in two studies reported here (Bruce, 1964; Cromer, 1970). It may be appropriate in future studies to look more closely at this relationship.

## REFERENCES

Al-Issa, I. The development of word definitions in children. *The Journal of Genetic Psychology*, 1969, **114**, 25–28.

Anglin, J. M. *The growth of word meaning*. Cambridge: M. I. T. Press, 1970.

Anisfeld, M., Barlow, J., & Frail, C. M. Distinctive features in the pluralization rules of English speakers. *Language and Speech*, 1968, **11**, 31–37.

Anisfeld, M., & Gordon, M. On the psychophonological structure of English inflectional rules. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 973–979.

Anisfeld, M., & Tucker, G. R. English pluralization rules of six-year-old children. *Child Development*, 1967, **38**, 1201–1217.

Asch, S. E., & Nerlove, H. The development of double function terms in children: An exploratory investigation. In B. Kaplan & S. Wapner (Eds.), *Perspectives in psychological theory: Essays in honor of Heinz Werner*. New York: International Universities Press, 1960.

Bradshaw, W. L., & Anderson, H. E., Jr. Developmental study of the meaning of adverbial modifiers. *Journal of Educational Psychology*, 1968, **59**, 111–118.

Berko, J. The child's learning of English morphology. *Word*, 1958, **14**, 150–177.

Bever, T. G. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley, 1970.

Bruce, D. J. The analysis of word sounds by young children. *British Journal of Educational Psychology*, 1964, **34**, 158–159.

Bryant, B., & Anisfeld, M. Feedback versus no-feedback in testing children's knowledge of English pluralization rules. *Journal of Experimental Child Psychology*, 1969, **8**, 250–255.

Chai, D. T. *Communication of pronominal referents in ambiguous English sentences for children and adults*. Unpublished doctoral dissertation, University of Michigan, 1967.

Chomsky, C. *The acquisition of syntax in children from 5 to 10*. Cambridge: M. I. T. Press, 1969.

Chomsky, N. Comments for project literacy meeting. (Project Literacy Rep., No. 2) Ithaca, N. Y.: Cornell University, 1964. (Mimeo) (Also available from ERIC Document No. ED 010 308.)

Chomsky, N., & Halle, M. *The sound pattern of English*. New York: Harper & Row, 1968.

Cromer, R. F. "Children are nice to understand": Surface structure clues for the recovery of deep structure. *British Journal of Psychology*, 1970, **61**, 397–408.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. Speech perception in infants. *Science*, 1971, **171**, 303–306.

Ervin, S. M. Imitation and structural change in children's language. In E. H. Lenneberg (Ed.), *New directions in the study of language*. Cambridge: M. I. T. Press, 1964.

Ervin, S. M., & Foster, G. The development of meaning in children's descriptive terms. *Journal of Abnormal and Social Psychology*, 1960, **61**, 271–275.

Fraisse, P., & Vautrey, P. The influence of age, sex, and specialized training on the vertical-horizontal illusion. *The Quarterly Journal of Experimental Psychology*, 1956, **8**, 114–120.

Fraser, C., Bellugi, U., & Brown, R. Control of grammar in imitation, comprehension, and production. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 121–135.

Gaer, E. P. Children's understanding and production of sentences. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 289–294.

Goldman-Eisler, F., & Cohen, M. Is N, P, and NP difficulty a valid criterion of transformational operations? *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 161–166.

Goodglass, H., Gleason, J. G., & Hyde, M. R. Some dimensions of auditory language comprehension in aphasia. *Journal of Speech and Hearing Research*, 1970, **13**, 595–606.

Halle, M., & Keyser, S. *English stress: Its form, its growth, and its role in verse*. New York: Harper and Row, 1971.

Hayhurst, H. Some errors of young children in producing passive sentences. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 634–639.

Hunt, K. W. *Grammatical structures written at three grade levels*. (Research Rep. No. 3) Champaign, Ill.: National Council of Teachers of English, 1965.

Huttenlocher, J., & Strauss, S. Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 300–304.

Huttenlocher, J., Eisenberg, K., & Strauss, S. Comprehension: Relation between perceived actor and logical subject. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 527–530.

Jacobson, R. *Child language, aphasia and phonological universals*. The Hague: Mouton, 1968.

Katz, E. W., & Brent, S. B. Understanding connectives. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 501–509.

Kessel, F. S. The role of syntax in children's comprehension from ages six to twelve. *Monographs of the Society for Research in Child Development*, 1970, **35**(6, Whole No. 139).

Liberman, A. M. General discussion. In F. Smith & G. A. Miller (Eds.), *The genesis of language*. Cambridge: M. I. T. Press, 1966.

Loban, W. D. *The language of elementary school children*. (Research Rep. No. 1) Champaign, Ill.: National Council of Teachers of English, 1963.

Loban, W. D. *Problems in oral English*. (Research Rep. No. 5), Champaign, Ill.: National Council of Teachers of English, 1966.

Lumsden, E. A., Jr., & Poteat, B. W. S. The salience of the vertical dimension in the concept of "bigger" in five and six year olds. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 404–408.

MacNeilage, P. F. Motor control of serial ordering of speech. *Psychological Review*, 1970, **77**, 182–196.

McNeill, D. *The acquisition of language: The study of developmental psycholinguistics*. New York: Harper & Row, 1970.

Menyuk, P. Syntactic structures in the language of children. *Child Development*, 1963, **34**, 407–422.

Menyuk, P. Syntactic rules used by children from preschool through first grade. *Child Development*, 1964 **35**, 533–546.

Menyuk, P. The role of distinctive features in children's acquisition of phonology. *Journal of Speech and Hearing Research*, 1968, **11**, 138–146.

Menyuk, P. *The acquisition and development of language*. Englewood Cliffs, N. J.: Prentice-Hall, 1971.

Natalicio, D. S., & Natalicio, L. F. S. "The child's learning of English morphology" revisited. *Language Learning*, 1969, **19**, 205–215.

Neimark, E. D. Development of comprehension of logical connectives: Understanding of "or." *Psychonomic Science*, 1970, **21**, 217–219.

Neimark, E. D., & Slotnick, N. S. Development of the understanding of logical connectives. *Journal of Educational Psychology*, 1970, **61**, 451–460.

O'Donnell, R. C., Griffin, W. J., & Norris, R. C. *Syntax of kindergarten and elementary school children: A transformational analysis*. (Research Rep. No. 8) Champaign, Ill.: National Council of Teachers of English, 1967.

Olds, H. F., Jr. *An experimental study of syntactic factors influencing children's comprehension of certain complex relationships*. (Center for Research and Development of Educational Difficulties Rep. No. 4), Cambridge, Mass.: Harvard University Press, 1968.

Olson, D. R. Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, 1970, **77**, 257–273.

Piaget, J. Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology*. New York: Wiley, 1970.

Poole, I. Genetic development of articulation of consonant sounds in speech. *Elementary English Review*, 1934, **11**, 158–161.

Savin, H. B., & Bever, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 295–302.

Shine, D., & Walsh, J. F. Developmental trends in the use of logical connectives. *Psychonomic Science*, 1971, **23**, 171–172.

Sinclair-de Zwart, H. Developmental psycholinguistics. In D. Elkind & J. H. Flavell (Eds.), *Studies in cognitive development: Essays in honor of Jean Piaget*. Oxford: Oxford University Press, 1969.

SLOBIN, D. I. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 219–227.

SLOBIN, D. I. Imitation and grammatical development in children. In N. S. Endler, L. R. Boulter, & Osser, H. (Eds.), *Contemporary issues in developmental psychology*. New York: Holt, Rinehart & Winston, 1968.

SLOBIN, D. I. *The ontogenesis of grammar: Facts & theories*. New York: Academic Press, 1971. (a)

SLOBIN, D. I. Developmental psycholinguistics. In W. O. Dingwell (Ed.), *A survey of linguistic science*. College Park, Md.: Linguistics Program, University of Maryland, 1971. (b)

SLOBIN, D. I., & WELCH, C. A. Elicited imitation as a research tool in developmental psycholinguistics. In C. A. Ferguson & D. I. Slobin (Eds.), *Readings on child language acquisition*. New York: Holt, Rinehart & Winston, in press.

SNOW, K. A comparative study of sound substitutions used by "normal" first grade children. *Speech Monographs*, 1964, **31**, 135–142.

TEMPLIN, M. C. *Certain language skills in children: Their development and interrelations*. (Monographs Series No. 26) Minneapolis: University of Minnesota, Institute of Child Welfare, 1957.

TURNER, E. A., & ROMMETVEIT, R. The acquisition of sentence voice and reversibility. *Child Development*, 1967, **38**, 649–660. (a)

TURNER, E. A., & ROMMETVEIT, R. Experimental manipulation of the production of active and passive voice in children. *Language and Speech*, 1967, **10**, 169–180. (b)

WARREN, R. M., & WARREN, R. P. A comparison of speech perception in childhood, maturity, and old age by means of the verbal transformation effect. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 142–146.

WELLMAN, B. L., CASE, I. M., MENGERT, I. G., & BRADBURY, D. Speech sounds of young children. *University of Iowa Studies in Child Welfare*, 1931, **5**, No. 2.

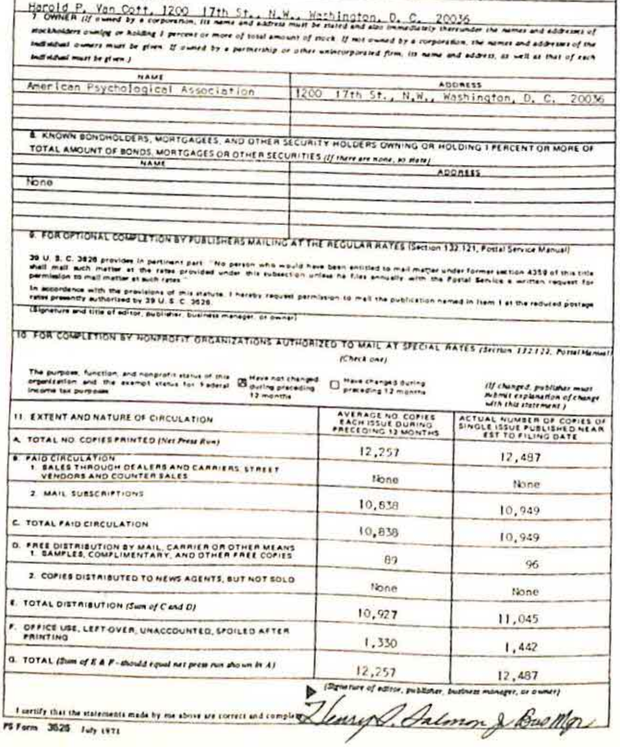
WHITE, S. H. Evidence for a hierarchical arrangement of learning processes. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior. Vol. 2*. New York: Academic Press, 1965.

WOLMAN, R. N., & BAKER, E. N. A developmental study of word definitions. *Journal of Genetic Psychology*, 1965, **107**, 159–166.

WRIGHT, P. Transformations and the understanding of sentences. *Language and Speech*, 1969, **12**, 156–166.

(Received July 6, 1971)

---



Page 1

U. S. POSTAL SERVICE

**STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION**

*(Act of August 12, 1970: Section 3685. Title 39. United States Code)*

SEE INSTRUCTIONS ON PAGE 2 (REVERSE)

1. TITLE OF PUBLICATION: PSYCHOLOGICAL BULLETIN

2. DATE OF FILING: 9/27/72

3. FREQUENCY OF ISSUE: MONTHLY

4. LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, city, county, state, ZIP code) (Not printers)*: Prince & Lemon Sts., Lancaster, Pa. 17604

5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*: 1200 17th Street, N.W., Washington, D. C. 20036

6. NAMES AND ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and address)*: American Psychological Association, 1200 17th St., N.W., Washington, D. C. 20036

EDITOR *(Name and address)*: James Deese, University of Virginia, Charlottesville, Virginia 22903

MANAGING EDITOR *(Name and address)*: Harold P. Van Cott, 1200 17th St., N.W., Washington, D. C. 20036

7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)*

| NAME | ADDRESS |
|---|---|
| American Psychological Association | 1200 17th St., N.W., Washington, D. C. 20036 |

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

| NAME | ADDRESS |
|---|---|
| None | |

9. FOR OPTIONAL COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES (Section 132.121, Postal Service Manual)

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in Item 1 at the reduced postage rates presently authorized by 39 U. S. C. 3626.

(Signature and title of editor, publisher, business manager, or owner)

10. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 132.122, Postal Manual)*

*(Check one)*

The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes ☒ Have not changed during preceding 12 months ☐ Have changed during preceding 12 months *(If changed, publisher must submit explanation of change with this statement.)*

| 11. EXTENT AND NATURE OF CIRCULATION | AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS | ACTUAL NUMBER OF COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE |
|---|---|---|
| A. TOTAL NO. COPIES PRINTED *(Net Press Run)* | 12,257 | 12,487 |
| B. PAID CIRCULATION<br>1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES | None | None |
| 2. MAIL SUBSCRIPTIONS | 10,838 | 10,949 |
| C. TOTAL PAID CIRCULATION | 10,838 | 10,949 |
| D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS<br>1. SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES | 89 | 96 |
| 2. COPIES DISTRIBUTED TO NEWS AGENTS, BUT NOT SOLD | None | None |
| E. TOTAL DISTRIBUTION *(Sum of C and D)* | 10,927 | 11,045 |
| F. OFFICE USE, LEFT-OVER, UNACCOUNTED, SPOILED AFTER PRINTING | 1,330 | 1,442 |
| G. TOTAL *(Sum of E & F—should equal net press run shown in A)* | 12,257 | 12,487 |

I certify that the statements made by me above are correct and complete. (Signature of editor, publisher, business manager, or owner) Henry S. Salmon Jr. Bus. Mgr.

PS Form 3526 July 1971

*Psychological Bulletin*
1972, Vol. 78, No. 6, 429 432

# NONPARAMETRIC INDEXES OF SENSITIVITY AND RESPONSE BIAS[1]

JOHN T. E. RICHARDSON[2]

*University of California, San Diego*

Dissatisfaction with theoretically based indexes of sensitivity and bias in detection and recognition experiments has motivated research into nonparametric indexes. Various suggestions from past work are presented and discussed with regard to their independence of particular theories. A particularly satisfactory measure of sensitivity is the area under the receiver operating characteristic curve. While no measure of bias is independent of sensitivity, the false-alarm rate can itself be used when experiments fail to reveal a change in the latter.

Several authors have recently pointed to difficulties in testing the assumptions made by theories of detection and recognition (Green & Swets, 1966, p. 78; Krantz, 1969; Lockhart & Murdock, 1970; Wickelgren, 1968). If we wish to employ the data such theories are supposed to explain in the case of other theoretical issues which might in practice be resolved, it is necessary to develop measures of sensitivity and bias that are independent of particular theories. This paper considers the various suggestions that have been made toward nonparametric indexes of detection and recognition, and examine their claims to theory independence.

## *Experimental Procedures*

The need to deal with the observer's decision criterion is practically eliminated by the use of forced-choice tasks, where the subject's sensitivity is measured directly by the proportion of correct responses he achieves (see Green & Swets, 1966, pp. 107–114). Brown (1965; Brown & Routh, 1970) has extended this procedure in a recognition situation by making subjects order the alternatives according to their subjective "oldness." The forced-choice procedure thus excludes, or at least minimizes, the role of motivational and response processes. This may be advantageous in situations where there is little theoretical motivation to study response bias, but may be unhelpful in other situations where response processes play a significant role, such as vigilance tasks. A further disadvantage with a multialternative forced-choice task is that the complexity of the decision might tend to increase the amount of error in the subject's response: this is particularly possible with the complex response task involved in Brown's procedure.

The chief alternative to the forced-choice procedure is the yes–no task, where subjects are asked to judge each of a set of target events and distracting events as to whether it is a target. The conventional representation of the subject's performance is the receiver operating characteristic (ROC), which plots the hit rate against the false-alarm rate, both rates being assumed to be functions of both sensitivity and response bias. Each point on the ROC curve represents a certain response bias on the part of the subject, and since the ROC curve is defined as the locus of all possible response biases, the whole curve is independent of any particular bias.

While the fundamental detection problem is the yes–no task, this procedure yields only one point on the ROC curve for a given experimental situation. It may of course be repeated using different payoffs, but a more efficient method of obtaining recognition scores is the use of a rating scale, where subjects are required to record the confidence with which they make their yes–no judgment; by a familiar method, $(n-1)$ scores can be obtained with an $n$-point rating scale for a single set of responses.

---

[1] This paper was supported by Public Health Service Grant MH 15828 from the National Institute of Mental Health to the Center for Human Information Processing, Department of Psychology, University of California, San Diego.

[2] The author would like to express gratitude to A. D. Baddeley, D. A. Norman, and D. E. Rumelhart for valuable discussions; and to D. M. Green and P. T. Smith for criticisms of an earlier draft of this paper.

Requests for reprints should be sent to John T. E. Richardson, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QY, Sussex, England.

While the rating procedure is more efficient and yields more stable results than the basic yes–no design (Swets, Tanner, & Birdsall, 1961, p. 329), no conclusive evidence has been produced to show that the rating procedure does not change the resulting ROC curves (for a discussion of relevant experiments, see Coombs, Dawes, & Tversky, 1970, pp. 197–198). This is essentially an assumption that must be tested for each situation studied.

For the remainder of this paper, attention is confined to the yes–no and rating procedures, examining attempts to express sensitivity and response bias in terms of the empirical variables of hit rate and false-alarm rate.

### *Sensitivity Indexes*

Weak assumptions apparently necessary for any sensible analysis of data from yes–no and rating tasks are that the ROC curve is continuous and that it includes a point whose hit rate is unity and another whose false-alarm rate is zero. Norman (1964, p. 245) assumed the ROC curve to be monotonically increasing with nonincreasing slope, and showed how the unit square could then be divided up for a given data point, such that another point could be said to represent superior or inferior performance, or could be said not to reflect determinably different performance. However, his method has several disadvantages: first, there are no obvious statistical tests that might be used with Norman's (purely graphical) method; second, computation for more than one or two data points would be very arduous; third, it offers no single numerical measure of sensitivity.

An important step was the demonstration by Green (1964) that a wide variety of theoretical models of detection processes agreed in predicting that the area under the ROC curve is equal to the proportion of correct responses attainable in a two-alternative, forced-choice experiment. Green and Swets (1966, pp. 81–84) discovered that Luce's (1959) choice theory does not predict this relationship, but they showed that if any assumptions about the underlying density function of the effects of stimulation are made, then the required equality holds. For this reason, they concluded that the area under the ROC curve provides a convenient and simple index of the detectability of the target, both in detection and in recognition experiments (Green & Swets, 1966, p. 50, pp. 344–345; see also Banks, 1970, p. 85).

Pollack and Norman (1964) pointed out that Norman's (1964) graphic method placed upper and lower bounds on the area under the ROC curve, and they used the mean of the two bounds as an estimate of the area, giving the isosensitivity curves predicted by this analysis, that is, the locus of points on the unit square which yield equivalent areas. Unfortunately, the analysis does indeed *predict* such curves, which (contrary to Pollack and Norman's assertions) diverge in important ways from those predicted by some theories, most particularly threshold theories. Their measure is not therefore theory independent, but makes highly specific predictions about the shape of the ROC curve. The fact that it has no theoretical underpinning only means that besides being too specific it is also conceptually degenerate. A particular theory, such as signal detection theory, would also be specific in its predictions, but would at least have a definite (indeed, a rich) conceptual content. Similar remarks apply to the measure suggested by Young (1971).

Pollack, Norman, and Galanter (1964) combined the employment of the area under the ROC curve with the assumption of the validity of the rating-scale procedure. This enabled them to generate simultaneously several points on the ROC curve, and thus to get a good estimate of the area under the curve using the trapezoidal rule. While they do not explicitly make the assumption that the ROC curve is monotonically increasing, it is dictated by the rating procedure itself, which cumulates data from one rating category to the next to yield successive points on the ROC curve. However, all of the theories described by Green and Swets (1966) assume a monotonically increasing ROC curve, and it is hard to conceive what a curve that lacked this property could represent, except perhaps an ROC curve derived from very noisy data.

If we accept that the rating procedure yields the same ROC curve as the yes–no procedure (see above), we may proceed to estimate the area by this method. The measure will be bounded by the values 0 and 1, and will tend to be above .5. Beyond this, however, its distribution is not determined at all by the analysis which produced it, only by the experimental situation (Green & Swets, 1966, p. 50), and will be generally unknown. It will therefore be possible only to use the ordinal properties of data in statistical analyses. But although choice theory yields predictions of sensitivity different from the area measure, the two do seem to be ordinally correlated (Green & Swets, 1966, p. 84). Insofar as choice theory was the only exception to Green's (1964) conclusion relating the area under the ROC curve to the proportion of correct responses in a two-alternative forced-choice test, we may conclude that Pollack et al.'s (1964) measure is truly independent of all theories of recognition and detection with respect to those of its properties which can actually be used in statistical analysis.

### *Response Bias*

Little work has been done toward finding a nonparametric estimate of subjects' biases. The only published paper is that of Hodos (1970), who attempted a graphical method along the lines of Pollack and Norman's. Again, however, a nonparametric method yields highly specific predictions. Hodos gave two families of isobias contours (each contour representing the locus of points having the same bias) that have highly specific properties and that differ considerably from those predicted by threshold theories. Further, if a data point occurs in the lower half of the unit square, Hodos' formulas yield a totally misleading result; such an event can occur due to chance variation if performance is low and is predicted by signal detection theory when the variances of the underlying signal and noise distributions are unequal.

Since there is no procedure analogous to the rating-scale method that could yield empirical isobias curves, it is unlikely that any bias estimate independent of sensitivity can be given. The best that can be done is to give a convenient measure for use when no change in sensitivity is found. Hodos' expressions are inadequate for the reasons given above, and because of their complicated formulation. But, more importantly, if the ROC curve is assumed to be monotonically increasing, then any suggested measure will be ordinally correlated with both the hit rate and the false-alarm rate. Both of these measures are of course given, and either might be used in preference.[3] Which is more convenient or successful will depend upon the particular experimental situation, although the hit rate is of necessity greater than the false-alarm rate at least most of the time, and so it may suffer from ceiling effects. One might therefore use the false-alarm rate as a measure of bias that is truly atheoretical.

### *Conclusion*

On the basis of the above discussion we can therefore recommend the use of the area under the ROC curve derived from the use of a rating scale as a measure of sensitivity independent of subjects' biases. As a measure of bias, one might use either the false-alarm rate or the hit rate, though both are dependent upon sensitivity and can only be used in comparisons where the latter is constant. Because the distribution of none of these measures is defined, only their ordinal properties may be used in statistical tests. This dictates the use of nonparametric statistics, but helps to ensure that the use of these measures is consistent with any possible theory of detection or recognition.

---

[3] Since one is once more limited to considering the ordinal properties of data. I am grateful to P. T. Smith for pointing this out.

## REFERENCES

BANKS, W. P. Signal detection theory and human memory. *Psychological Bulletin*, 1970, 74, 81–89.

BROWN, J. Multiple response evaluation of discrimination. *British Journal of Matematical and Statistical Psychology*, 1965, 18, 125–137.

BROWN, J., & ROUTH, D. A. Recognition assessed by *d'* and by a non-parametric alternative (the *A*-index) as a function of the number of choices. *Quarterly Journal of Experimental Psychology*, 1970, 22, 707–719.

COOMBS, C. H., DAWES, R. M., & TVERSKY, A. *Mathematical psychology: An elementary introduction*. Englewood Cliffs, N. J.: Prentice-Hall, 1970.

GREEN, D. M. General prediction relating yes–no and forced-choice results. *Journal of the Acoustical Society of America,* 1964, **36**, 1042.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

HODOS, W. Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin,* 1970, **74**, 351–354.

KRANTZ, D. H. Threshold theories of signal detection. *Psychological Review,* 1969, **76**, 308–324.

LOCKHART, R. S., & MURDOCK, B. B., JR. Memory and the theory of signal detection. *Psychological Bulletin,* 1970, **74**, 100–109.

LUCE, R. D. *Individual choice behavior.* New York: Wiley, 1959.

NORMAN, D. A. A comparison of data obtained with different false-alarm rates. *Psychological Review,* 1964, **71**, 243–246.

POLLACK, I., & NORMAN, D. A. A non-parametric analysis of recognition experiments. *Psychonomic Science,* 1964, **1**, 125–126.

POLLACK, I., NORMAN, D. A., & GALANTER, E. An efficient non-parametric analysis of recognition memory. *Psychonomic Science,* 1964, **1**, 327–328.

SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. Decision processes in perception. *Psychological Review,* 1961, **68**, 301–340.

WICKELGREN, W. A. Testing two-state theories with operating characteristics and a posteriori probabilities. *Psychological Bulletin,* 1968, **69**, 126–131.

YOUNG, S. J. A note on the relation between measures of the percent correct and the area under the ROC curve for verbal recognition studies. In E. J. Crothers et al., *Studies in mathematical learning theory and psycholinguistics.* (Publ. No. 4) Denver, Colo.: Computer Laboratory for Instruction in Psychological Research, University of Colorado, 1971.

(Received June 14, 1971)

---

## Manuscripts Accepted for Publication in the *Psychological Bulletin*

N-of-One and N-of-Two Research in Psychotherapy. John M. Gottman (Psychological Clinic, Psychology Building, Indiana University, Bloomington, Indiana 47401).

Further Thoughts on Thinking and Language. Hans G. Furth (Department of Psychology, Catholic University of America, Washington, D. C. 20017).

Criterion Shifts in Recall. Alice F. Healy (The Rockefeller University, New York, New York 10021) and Cindy Jones.

*Psychological Bulletin*
1972, Vol. 78, No. 6, 433-446

# DECISION THEORY ANALYSIS OF SOCIAL POWER[1]

WILLIAM E. POLLARD[2] AND TERENCE R. MITCHELL

*University of Washington*

The purpose of this paper is to discuss the use of decision theory as a framework for clarifying and organizing various conceptualizations of social power. The theory of decision making used in our analysis suggests that an actor's decision regarding the performance of some behavior is a function of (*a*) the actor's utility or evaluation of the consequences of the behavior and (*b*) the actor's subjective probabilities or expectations that the behavior will lead to these consequences. The decisions and behavior of actors in exerting and responding to social power are analyzed in this manner. A number of conceptual issues are discussed in these terms, and several theories of social power are examined in terms of concepts and distinctions derived from the decision theory analysis of social power.

The topic of social power has interested and intrigued numerous social scientists and social philosophers. "That some people have more power than others is one of the most palpable facts of human existence," wrote Robert Dahl (1957, p. 201) in his discussion of social power. "Because of this," he continued, "the concept of power is as ancient and ubiquitous as any that social theory can boast [p. 201]." Indeed, a number of writers have viewed the concept of power as occupying a central position in the social sciences. Bertrand Russell (1938) wrote, "I shall be concerned to prove that the fundamental concept in social science is Power, in the same sense in which Energy is the fundamental concept in physics [p. 9]." Social psychologists Dorwin Cartwright and Kenneth Clark have both argued that the study of power should be a central concern of social psychology (Cartwright, 1959a, p. 13; Clark, 1965, p. 16), and Clark, like Russell, stated that "A theory of power could serve not only as a unifying principle for social psychology but as *a unifying principle for social science as a whole* [p. 16]."

Despite the level of interest in the topic, the concept of power has yet to occupy a central position in the social sciences as envisioned by the above writers. Cartwright (1965), Kornberg and Perry (1966), Riker (1964), and Schopler (1965), in review articles, all pointed out a disconcerting heterogeneity in the power literature. There are a number of conceptualizations of social power that differ in emphasis and in scope, and the relationship of these different views to each other is not clear.

The purpose of this paper is to suggest an approach to the study of power based on decision theory. The essential idea is that the operation of power processes can be studied in terms of the decisions men make in exerting power and in responding to the exertion of power. Decisions made in response to power will be examined in the most detail since this has been the major focus of work in this area. This approach may be useful in clarifying certain issues in the analysis of social power and in providing a framework for integrating various theories of social power with each other and with more general theories of behavior.

## REVIEW OF CURRENT CONCEPTIONS OF SOCIAL POWER

This section briefly reviews the main points of a number of theories that have been used in social psychological analyses of power. The main points of the various theories are highlighted for later discussion of a decision theory analysis of power. More detailed reviews can be found in Collins and Raven (1969), Kornberg and Perry (1966), Schopler (1965), Shaw and Costanzo (1970), and Riker (1964).

Kornberg and Perry (1966) suggested that theories of power can be arranged along an ends–means continuum. Theories are classified according to the aspect of the power relationship that is dealt with in the most detail. On one end of the continuum are located those theories that focus on the means or process by which power is exercised, while on the other end are located those theories that focus on the end state or outcome of the power relationship. Of course, most writers discuss both process and outcome; but there does appear to be a real difference in emphasis. Of the theories discussed here, the field theories and the exchange theories are process theories, whereas the political science decision-making theories are outcome theories.

---

[1] This study was supported by Contract N00014-67-A-0103-0013, Advanced Research Projects Agency Order 454 (Fred. E. Fiedler, Principal Investigator).

[2] Requests for reprints should be sent to William E. Pollard, Organizational Research, University of Washington, Seattle, Washington 98105.

The actor attempting to exert influence is represented as A, and the other actor as B. To avoid confusion, citations that used other letters to symbolize the interaction of two actors are changed in accordance with this convention.

### *Field Theory*

This approach to the study of power is based on Lewin's field theory in which behavior is seen as a function of needs or tensions, conceptualized as force vectors. Although Lewin himself did not discuss power in great detail, his approach was elaborated by a number of psychologists at the University of Michigan in the 1950s. Since the approaches of different writers vary somewhat, we briefly review Lewin's work and then focus on Cartwright's analysis, which has been described in the most detail (Cartwright, 1959b; Cartwright & Zander, 1968).

Lewin (1951, p. 336) defined the power of A over B as the quotient of the maximum force that A can induce on B and the maximum resistance that B can offer. He pointed out, however, that he conceived of power as referring to a "possibility of inducing forces [p. 40]"; thus, the statement that A has power over B does not necessarily mean that A is actually inducing forces in B at that time.

Cartwright (1959b) defined power as follows: "The power of A over B with respect to a given change at a specified time equals the maximum strength of the resultant force which A can set up in that direction at that time [p. 193]." A's act is seen as creating in B both a force to compy and a force to resist; thus the resultant force (power) is given by the difference between these two. He pointed out that in Lewin's definition the force A can "induce" on B and the "resistance" that B can offer are terms that can be interpreted in a number of different ways. However, he stated that if in interpreting these terms we consider only forces activated by A, then the major difference between Lewin's and Cartwright's definitions is that Lewin defines power as a ratio of these forces while Cartwright defines power as a difference.

Whether the change in some state of B actually occurs is a function of the sum of all the forces operating on B at that time, and the outcome depends on whether the force set up by A is large enough to overcome opposition forces. A is said to have control over B's behavior when the change actually occurs. In these terms, A can have power over B without bringing about a behavioral change since B's actual behavior is a function of forces induced by other actors and the impersonal environment, in addition to forces created by A. Thus, A can have power by overcoming the resistance to him. To have control, however, he must overcome not only resistance forces but all the opposition forces as well. One should note that power is viewed here as a possibility for inducing forces—it refers to the maximum effect A could have if he actually decided to use his power.

Two important factors that determine A's power are the resources of A and the needs or "motive bases" of B. The idea is that "an act of A must 'tap' a motive base in order for it to activate a force [Cartwright, 1959b, p. 205]." In other words, A is able to create a force in B through the use of resources that can affect B's need satisfaction. French and Raven (1959) provided a detailed discussion of some of the different motive bases involved.

### *Exchange Theory*

The exchange theories provide an economic-like analysis of the interaction between actors. The exchange theorists view this interaction in terms of the rewards and costs the actors mediate for each other, and focus on the exchange aspect of the mutual dispensation of rewards and punishments or costs. The general assumption is that individuals behave so as to maximize the difference between the rewards and costs they experience. In these terms, A has power over B and can influence his behavior to the extent that he can determine or control the rewards and costs B experiences, Thibaut and Kelley (1959) view interaction in terms of each actor selecting and performing some behavior, or sequence of behaviors, from his behavioral repertoire. This combination of each actor's behaviors is called an outcome for the actors involved. Thibaut and Kelley illustrated this in matrix form where A's behavioral repertoire is listed along the top of the matrix, and B's along the side. The outcomes of all possible combinations of behaviors can be seen in the cells at the intersections of the columns and rows representing A's and B's behaviors, respectively. The rewards and costs that A and B would experience from the various outcomes are displayed in each cell. The rewards and costs that an actor would experience can come from both the other actor and intrinsic satisfactions and dissatisfactions associated with the behavior. These rewards and costs can theoretically be combined so as to give some number representing the goodness or value of the outcome for the actor. Thibaut and Kelley pointed out that the rewards and costs entered in the cells are seen as an objective statement of the consequences of the outcomes. In other words, they represent what the actors will actually experience and not necessarily what they perceive or anticipate that they will experience. Thibaut and Kelley argued that over continued interactions the actors' perceptions may become more accurate and come to resemble the objective matrix. However, it is important to recognize that the matrix does not necessarily represent the situation as the actors see it.

Certain outcomes will, of course, have greater value than others for an actor. However, an actor's attainment of certain outcomes is dependent upon the other actor's choice of behaviors. Thus Thibaut and Kelley conceptualized A's power over B as A's ability to influence the outcomes B experiences. The power of A over B is a function of the range of outcome values that A can determine for B. For example, if B valued all outcomes equally, A would have no power, for his choice of behavior could have no effect on the outcome value B would experience. However, if B valued one outcome more than the others, A would have power over B through his ability to choose or not choose the behavior from his repertoire that would provide that outcome, and his power would be a direct function of the range or difference in outcome values he could determine for B. The upper limit to this range is the best outcome B could obtain in interacting with A; however, the lower limit depends on the value of outcomes B could obtain in some other relationship. If B's outcomes are poorer than he could get elsewhere, he will be likely to leave the relationship. This sets a lower limit on this range. Thibaut and Kelley also pointed out that A may not be able to use the full extent of his power as the performance of certain necessary behaviors may be quite costly to him. Thus, power is usable only to the degree that it does not penalize its possessor. Therefore, B's counterpower can limit A's power. Their approach can be expanded to deal with the interaction among a number of actors by adding new dimensions to the outcome matrix for the additional actors. However, this makes the analysis quite complicated, so most of their discussion of power concerns only two actors.

Emerson's (1962) theory is similar to Thibaut and Kelley's theory. His main focus, however, is more on the balance of power and changes toward balance. He argued that "power resides implicitly in the other's dependency [p. 194]." The dependence of B upon A is directly proportional to the motivational investment (value) B has in the goals mediated by A, and inversely proportional to the availability of these goals outside the A–B relationship. Power is defined as "the amount of resistance on the part of B which can be potentially overcome by A [p. 194]," and it is argued that the power of A over B is equal to, and is based upon, the dependence of B upon A.

A balance of power, in which the power each actor has over the other is equal, does not mean that power is somehow cancelled out. Emerson (1962) defined cohesion as the level of mutual dependency in the relationship; thus, power can be balanced at various levels of cohesion. To view power as cancelled out in a state of balance is to ignore the consequences for the cohesion of the relationship. Emerson argued that there is a tendency for power relationships to move toward a state of balance. He provided a detailed discussion of the processes involved and analyzed norms, roles, authority, and status in these terms. In later papers, Emerson (1969; in press, a; in press, b) has expressed these ideas in reinforcement theory terms and has gone into the analysis of group processes in more detail.

Although these theories use different terms and have somewhat different emphases, they are generally similar in that they argue that A has power to influence B's behavior through his ability to manipulate the rewards and costs, or positive and negative values, that B experiences in exchange for the desired behavior on the part of B. Blau (1964), Karlsson (1962), and Kuhn (1963) have provided similar analyses of social power.

### *Political Science Decision-Making Theory*

Although this approach has a political science orientation, it is often referred to in the social psychological literature and is relevant to social psychological analyses of power. It should not be confused with decision theory as discussed in this paper, for although the early work appeared to be going in the same direction, the two theories have since diverged.

March (1955) argued that power (he used the term "influence") should be studied in the context of decision making. He stated that "influence is to the study of decision making what force is to the study of motion—a generic explanation of the basic observable phenomena [p. 432]." In other words, the power of an actor can be examined in terms of the effects he has on the decisions or choices of other actors. He then defined influence as "that which induces behavior on the part of the individual at time $t_1$ different from that which might be predicted on the basis of a knowledge of the individual organism at time $t_o$ [p. 438]," where behavior means any change in the state of the organism. Thus, "if the individual deviates from the predicted path of behavior, influence has occurred and, specifically . . . it is influence which has induced the change [p. 435]." March provided some discussion of influence in terms of changes in the values and expectations of actor B and the consequent effect on his decision. However, in later papers by him and others in this area, this psychological aspect of the theory is deemphasized, and the major emphasis is placed on the assessment of influence (or power) in terms of his definition.

Dahl (1957) elaborated on March's definition. He started with an intuitive definition of power, which is, "A has power over B to the extent that he can get B to do something B would not otherwise do [p. 202–203]." He worked from this idea to his view that the power of A over B is a function of the difference between the probability of B performing a given behavior after A has made some intervention and the probability of B performing this behavior without A's intervention. The amount of power is expressed as the difference between these two probabilities. Harsanyi (1962a, 1962b) has extended this analysis; however, his theory is mentioned below since it is related to the decision theory approaches.

As mentioned earlier, March and Dahl define power in terms of the outcome of the interaction, in contrast to the field theorists and exchange theorists who define power in terms of the process involved.

### *Summary*

We have presented several different types of theories that have been advanced to deal with the phenomenon of social power. Riker (1964), in his review article, pointed out the unusual fact that while these theories are all quite reasonable by themselves, they cannot be precisely translated into each others' terms. Yet it is difficult to find direct contradictions among them. The empirical literature is not very helpful here, for as Schopler (1965) pointed out:

> The major theoretical analyses of power typically have little data uniquely related to them, nor have they generated anything like a distinctive and coherent set of testable issues. The empirical research, on the other hand, although often acknowledging a particular theoretical parentage, is ordinarily tied to that systematic position in a tenuous manner [p. 179].

We feel that the decision theory analysis discussed below may be useful in providing a framework for clarifying some of the similarities and differences in the various definitions of social power. In the final section of the paper, we briefly examine various definitions of social power in terms of certain concepts and distinctions developed in our analysis.

## DECISION THEORY

The purpose of this section is to present a brief summary of the basic ideas from decision theory that are used in the discussion of power. In addition, there is a short review of how these ideas have appeared in a number of psychological theories. Finally, the appearance of decision theory concepts in the social power literature is discussed.

The basic ideas of decision theory, according to Edwards (1954), have their roots in the philosophical economic writings of Jeremy Bentham and James Mill, who maintained that the behavior of men is guided by the desire to obtain pleasure and avoid pain. Actions can be assessed in terms of the pleasurable or painful properties of the consequences, or outcomes, of the actions. These properties are known as the utilities of the outcomes, with pleasure seen as positive utility and pain seen as negative utility. The argument is that men choose among actions those with the maximum overall utility. This *felicific calculus* was later put in mathematical form and has, with refinement, come to form a basis for much economic and psychological thinking about human choice and decision making.

In some situations the actor may know exactly what the outcomes of the various courses of action will be. The actor's problem here is to select the course of action with the largest overall utility. This is called a riskless decision. In other situations, however, the actor may have to make a decision where the occurrence of certain outcomes as a consequence of the courses of action is a matter of probability. This is called a risky decision. The concept of subjective expected utility (SEU) is important here. The actor's SEU for an action is found by multiplying the actor's utility for the various outcomes of the action by the probability the actor associates with their occurrence, and summing across outcomes. Thus,

$$\mathrm{SEU} = \sum_{i=1}^{n} \psi_i u_i$$

with

$$\sum_{i=1}^{n} \psi_i = 1.0$$

where $n$ is the number of outcomes, $\psi$ is the actor's subjective probability that the behavior will lead to the outcome, and $u$ is the utility the actor attaches to the outcome. The requirement that the probabilities sum to one implies that the $n$ outcomes are independent and that all possible outcomes under consideration by the actor are taken into account. When considering risky decisions, the maximization assumption becomes one of maximization of subjective expected utility. Therefore, the decision involves comparing the SEUs for the various courses of action and selecting the course of action with the largest SEU.

One should note two points regarding this analysis. First, we are dealing with the actor's perception of the situation—the focus is on the actor's subjective evaluations or utilities of the outcomes and the actor's subjective probability estimates regarding the occurrence of the outcomes. Thus, the actor may hold socially deviant utilities or objectively inaccurate probabilities and yet perform as the SEU analysis would predict, given his utilities and subjective probabilities. Second, this type of analysis is often criticized for allegedly portraying man as an avaricious miser out to maximize his satisfaction at everyone else's expense. Actually, there is nothing in our analysis to prevent the individual from attaching positive utility to the welfare and happiness of others and behaving accordingly.

An important issue here is the question of how well predictions based on the SEU analysis correspond with the actual behavior of individuals. It is obvious that humans do not have unlimited information-processing capacities and this, of course, puts some limitation on the SEU analysis as a descriptive model of decision making. Nevertheless, there is evidence to suggest that individuals do, indeed, perform as the analysis predicts (see reviews by Becker & McClintock, 1967; Edwards, 1961; Peterson & Beach, 1967; Taylor, 1965). Additional support for this type of analysis is provided in the next section, where it is argued that the basic ideas of the SEU analysis correspond with the basic ideas in a number of psychological theories.

### *Related Approaches*

Theories similar to the decision-making theory discussed above have appeared independently in a number of areas of psychology. Although there are distinct differences in language and emphasis, the underlying idea is that behavior is a function of the probability or degree to which the behavior leads to various outcomes or consequences, and the value or utility of these outcomes. Mitchell and Biglan (1971) have reviewed a number of theories in the areas of verbal conditioning (Dulany, 1961, 1962, 1968), attitudes (Fishbein, 1963, 1967; Peak, 1955; Rosenberg, 1956), and worker motivation (Gailbraith & Cummings, 1967; Georgopoulos, Mahoney, & Jones, 1957; Graen, 1969; Hackman & Porter, 1968; Lawler, 1968; Lawler & Porter, 1967; Vroom, 1964) where this approach has been used with some success. Lawler (1971) has provided a similar review. Barclay (1971) has reviewed theories in a number of areas of psychology and has concluded that the learning theories of Hull (1943, 1952) and Tolman (1932, 1948), the personality theories of Rotter (1954) and Atkinson (1964), and Rosenberg's (1956) attitude theory can be interpreted in the decision theory terms discussed above. Feather (1959) and Scheibe (1970) have also pointed out a number of convergences among these theories. Thus, the fact that the same basic ideas have appeared independently in a number of theories lends support to this type of analysis. In addition, successful research in a number of these areas, along with the decision theory research mentioned above, suggests that the variables involved can be fruitfully measured and used in research. Finally, these underlying similarities imply that a decision theory analysis of power could be tied in with theories in a number of areas of psychology.

### *Relationship to Theories of Social Power*

It is possible to look at the theories of social power discussed above in decision theory terms since the process theories and the early form of March's political science decision-making theory all suggest that B's response to A's influence attempt is determined by A's ability to affect B's need satisfaction or provide rewarding and/or punishing outcomes for B contingent upon B's compliance or noncompliance. This is related to the decision theory idea that an actor's choice among alternative behaviors is a function of the utility of the outcomes of the behaviors for the actor. If we look at the theories of social power in these terms, it can be seen that they imply a riskless decision-making process. We feel that power relations in real-life settings are better viewed in terms of risky decisions. In other words, we are arguing that B's subjective probabilities that various outcomes will occur as a consequence of his compliance or noncompliance are important factors in determining his behavior and should consequently be included in a theory of social power.

The concept of subjective probability is not totally foreign to these theories since concepts equivalent to subjective probability can be found in the general theories of behavior on which the theories of social power are based. However, the concept of subjective probability has generally not been incorporated formally into the central arguments of the various theories of social power.

The concept of subjective probability has appeared in a number of places in the field theory literature. Lewin made a number of references to subjective probability in his field theory analysis of behavior; and his theory of level of aspiration (Lewin, Dembo, Festinger, & Sears, 1965) is formally equivalent to the SEU analysis. He did not, however, incorporate these ideas into his analysis of social power. Cartwright (1965) and French and Raven (1959) also made reference to subjective probabilities. Cartwright stated, "in general, the strength of force induced by an act of A will be greater the greater B's subjective probability that compliance will result in a favorable outcome with respect to a motive base [Cartwright, 1965, p. 32]," and argued in the same paper that the field theory approach "[has] much in common with other theories of decision making which employ such concepts as alternatives, expectations (subjective probability) and utility [p. 25]." French, Morrison, and Levinger (1960) and Zipf (1960), in their field theory research on social power, have actually conceptualized some of the forces in terms of the product of the valence or utility of outcomes for compliance and noncompliance multiplied by the subjective probabilities of their occurrence.

As we mentioned above, the exchange theories of social power are based on learning theory and economic concepts of exchange. With respect to learning theory, Homans (1969), along with others mentioned in the preceding section, argued that the idea of subjective probability can be tied in with the concept of history of reinforcement. The role of the history of reinforcement has not been elaborated in great detail in exchange theory analyses of social power, although Emerson discussed history of reinforcement in his more recent papers. The economic ideas used in exchange theories are quite similar to decision theory ideas; however, subjective probabilities have not been explicitly incorporated in the analyses.

As for the political science decision-making theories, March (1955) suggested looking at power in terms of B's utilities and subjective probabilities, although subsequent analyses by March and Dahl have emphasized the outcome of interaction rather than the process. Harsanyi (1962a, 1962b), however, has provided a process analysis in an extension of Dahl's theory. He views the use of power in terms of a bargaining game in which A and B negotiate the probability with which B will perform X, and he looks at this process in terms of the actors' utilities and subjective probabilities. Nagel (1968) has modified certain aspects of Harsanyi's analysis and has then shown that an SEU analysis of social power can be useful in dealing with a number of theoretical problems discussed in the social power literature. In addition, he shows that this type of approach can provide a useful framework for interpreting and integrating various theories of social power.

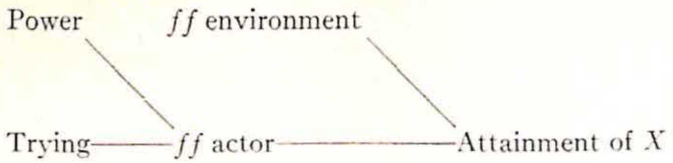
Our point here is that decision theory ideas are not inconsistent with the various theoretical analyses of social power. Indeed, there are differences in emphasis among the various theories, and each has its unique perspective on the analysis of social power. Yet we feel that a decision theory analysis of social power may serve as a basis for clarifying the underlying similarities and differences among the various theories and resolving certain issues regarding the analysis of social power.

## DECISION THEORY ANALYSIS OF SOCIAL POWER

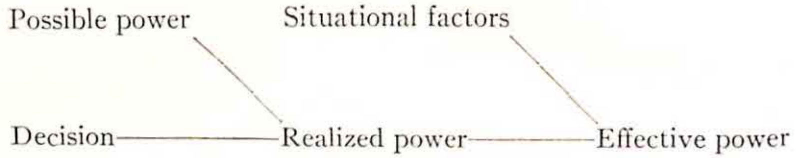
### *General Conception of Power*

In this section a rather general view of power is discussed that hopefully will provide an integrated view of a number of seemingly different conceptualizations of power. In the course of the discussion, a number of different terms are used to distinguish various aspects and types of power. This is not due to some fascination with coining terms, but rather is a result of an effort to distinguish a number of quite different phenomena that all have been referred to by the same term—power. The concept of power is first examined from the perspective of A and then from the perspective of B. Finally, certain aspects of the way we conceptualize the relationship between A's and B's perspectives are discussed. The term *actor* is used in our discussion to suggest that the analysis may apply to relations between groups as well as to individuals. It is important to recognize, however, that a complete theory of group behavior in power relations would require, in addition to the material discussed here, a more detailed analysis of group decision making.

Our starting point here is Kuhn's (1963, p. 317) general notion of power as the ability to satisfy wants or obtain goals. These goals may be social or nonsocial. Heider's analysis of an actor's goal attainment is useful to consider in this context (Heider, 1958, pp. 82–83). He views the ability of an actor to obtain a goal ($X$) as a function of the effective force of the actor (*ff* actor) and the effective force of the environment (*ff* environment). Whether or not actor A obtains his goal ($X$) depends on whether or not the effective force of A is great enough to overcome the opposing effective force of the environment. Heider argued that the effective force of A can be analyzed into a power factor and a motivational factor called trying. The power factor is a function of a number of things; however, ability is seen as being the most important. The motivational factor refers to A's intentions and effort exerted. Heider diagramed the relationship between these factors as follows:



We feel that these terms generally correspond to the different ways power has been conceptualized and measured. Thus, if the different conceptions of power are substituted in this diagram, some relationship among them can be seen:



A's possible power refers to the abilities and resources that A could use to obtain a goal. It refers to the effects that A could have in a given situation, but not necessarily to the effects that A will actually have in that situation. The term *possible power* refers to an objective assessment of what A could do, and this is not necessarily the same as A's own perceptions or other actors' perceptions regarding what A could do. A's decision involves the choice to actually use some of this possible power in a specific way to obtain some goal. This decision can be seen as a function of A's SEU for taking some goal-oriented action, in comparison to A's SEUs for alternative actions. Important factors are A's positive utilities associated with the goal, negative utilities associated with the expenditure of various resources, time, effort, etc., and A's probability estimates regarding the occurrence of these outcomes. Possible power plus the decision to use some of it in a specific way to obtain some goal determines A's realized power or the force that will actually be exerted by A in an effort to obtain the goal. A's effective power or his ability to actually obtain the goal is determined by A's realized power and situational factors that may affect goal attainment. These situational factors are often negative factors to be overcome; however, they can, in some situations, facilitate or support goal attainment. This framework combines various distinctions made by Dahlström (1966, p. 238), Nagel (1968, pp. 133–135), and Wrong (1968, pp. 677–681) in their discussions of social power.

### *Social Power*

Social power can be seen as a special case of this general conception of power—it refers to the situation where A's goal is to obtain certain behavior from another actor (B). A number of writers (Bachrach & Baratz, 1963; Harsanyi, 1962b; Riker, 1964) have pointed out, however, that there are two different types of social power. The first type involves the case where A obtains the desired behavior from B by influencing B's decision-making process. Thus A provides incentives, information, promises, threats, etc., in order to influence B's choice of actions or behaviors. The second type is more similar to nonsocial power in that A obtains his goal independently, and often in spite of B's decisions.

Following Harsanyi (1962b), this first type is referred to as incentive social power and the second type as independent social power. Independent social power is limited to certain types of goals; nevertheless, it is included frequently in discussions of social power and is discussed briefly in the next section before going on to the main topic of interest—incentive social power.

*Independent social power.* The most frequently discussed examples of this form of social power are physical force and the control of group decisions by coalitions. These phenomena are intuitively related to discussions of power; however, there is some confusion in the literature regarding their relationship to other theories of social power which are, for the most part, incentive theories. Their place in a theory of social power can best be seen in the context of a general framework including social and nonsocial power, for the form of power involved is social only in a limited sense. The actual power processes involved are similar to those of nonsocial power even though another actor is involved.

Force involves the actual physical control of A over B. It is not the threat of control—it involves actually physically overcoming the other actor. Examples can be seen in parents getting their children to bed by physically placing them there, the robber obtaining his victim's money by knocking him unconscious, etc. Van Doorn (1962) wrote, "force . . . implies that the power-subject is degraded to the status of an object. Since no choice is left to him, he loses practically all his human qualities and functions as an object of action [p. 25]." Goldhammer and Shils (1939) in fact classified force as nonsocial power.

Coalition theory in voting situations, as discussed by Gamson (1964) in a review article, and by Shapley and Shubik (1954) in their research on a legislative body, includes considerations of both incentive and independent social power. The actual formation of coalitions can be seen in terms of how the resources of actors (votes) influence the decisions of other actors in joining a coalition. However, the final decision of the group as a whole is seen simply as a matter of which coalition controls the votes. The coalition with the most votes is seen as automatically obtaining its goals, irrespective of what the minority wants or decides.

Thus, although these phenomena involve social interaction, B's decision making is not the crucial factor that it is when we consider incentive social power.

*Incentive social power*. This form of social power can be seen in the situation where A's goal is to get B to perform a certain act and A attempts to obtain this goal by influencing B's decision. Thus, A's power can be examined in terms of his effect on B's decision.

We can conceptualize B's decision process in the terms discussed above. In a given situation B has a number of discrete behaviors to choose from. The behavior desired by A is represented by $X$ and the set of remaining behaviors is represented by $\bar{X}$. Any specific behavior in $\bar{X}$ is referred to as $\bar{X}_i$. There are a number of possible outcomes for each of the behaviors. B's SEU for each of the behaviors can be determined by multiplying B's subjective probabilities that the behavior will lead to these outcomes times the utility B attaches to the outcomes. Thus we can obtain $\text{SEU}_X$, $\text{SEU}_{\bar{X}_1}$, $\text{SEU}_{X_2}$, . . . , $\text{SEU}_{\bar{X}_n}$.

The possible outcomes are viewed as having a number of components, some of which are determined by A and others which stem from situational factors such as other actors, the impersonal environment, intrinsic satisfactions or dissatisfactions that B experiences from performing certain types of behavior, etc. Thus B's SEUs are determined partly by A's manipulation of the outcomes and partly by the effects of situational factors on the outcomes. Consequently, B's SEUs for the behaviors can be broken down into two parts—one reflecting A's effects on the outcomes B anticipates experiencing and the other representing the effects of the situational (S) factors Therefore,

$$\text{SEU}_X = \text{SEU}_{X(\text{A})} + \text{SEU}_{X(\text{S})} \qquad [1]$$

and

$$\text{SEU}_{\bar{X}_i} = \text{SEU}_{\bar{X}_i(\text{A})} + \text{SEU}_{\bar{X}_i(\text{S})} \qquad [2]$$

According to the maximization assumption, B should choose the behavior with the maximum SEU. We are assuming here that there are some alternatives for which,

$$\text{SEU}_{\bar{X}_i(\text{S})} \geq \text{SEU}_{X(\text{S})}; \qquad [3]$$

that is, we are dealing with a situation in which, in the absence of the effects of A on the outcomes, B would choose some alternative other than $X$. If B would choose $X$ to begin with, there would be no need for A to try to influence B. In fact, only those $k$ alternatives that B would possibly choose in preference to $X$ are considered in our analysis.

Whether or not B will perform $X$ can be determined by comparing B's SEU for $X$ with B's SEUs for each of the alternative behaviors. If B is to perform $X$, the difference between $\text{SEU}_X$ and $\text{SEU}_{\bar{X}_i}$ should be positive for all the alternatives under consideration.

$$\text{SEU}_X - \text{SEU}_{\bar{X}_i} > 0 \qquad [4]$$

for $i = 1, 2, \ldots, k$. In other words, B's SEU for performing $X$ should be greater than his SEUs for any of the $k$ alternative behaviors.

In attempting to influence B, A's goal is to get B to perform $X$ and to not perform $\bar{X}_i$. Consequently, we would expect A to manipulate the outcomes that B anticipates in order to make B's $\mathrm{SEU}_X$ larger and B's $\mathrm{SEU}_{\bar{X}_i}$ smaller so that the differences in Expression 4 would be positive and B would perform $X$. Thus, a measure of the effect that A has on B's choice between $X$ and some $\bar{X}_i$ is the extent to which A contributes to the difference between $\mathrm{SEU}_X$ and $\mathrm{SEU}_{\bar{X}_i}$. Therefore, A's realized power over B with respect to a choice between $X$ and some $\bar{X}_i$ is given by

$$\mathrm{SEU}_{X(\mathrm{A})} - \mathrm{SEU}_{\bar{X}_i(\mathrm{A})}. \qquad [5]$$

Although it does not occupy a central place in our analysis, one could perhaps derive some index of the overall effect A has on B's decision by summing or averaging Expression 5 across alternatives. These procedures, however, raise a number of questions about the number and level of generality of the alternatives that we do not have space to deal with here.

Whether or not B will actually decide to perform $X$ is, as was discussed above, a function of Expression 4. If we substitute Equations 1 and 2 in Expression 4 and rearrange, we get:

$$(\mathrm{SEU}_{X(\mathrm{A})} - \mathrm{SEU}_{\bar{X}_i(\mathrm{A})}) + (\mathrm{SEU}_{X(\mathrm{S})} - \mathrm{SEU}_{\bar{X}_i(\mathrm{S})}) > 0 \qquad [6]$$

for $i = 1, 2, \ldots, k$. Actually, the question regarding what B will do can be reduced to a comparison between $\mathrm{SEU}_X$ and the *maximum* SEU of the alternatives ($\mathrm{SEU}_{\bar{X}\,\max}$). If $\mathrm{SEU}_X$ is larger, B should perform $X$; if $\mathrm{SEU}_{\bar{X}\,\max}$ is larger, B should perform $\bar{X}$ max. Thus, for this comparison, Expression 6 becomes

$$(\mathrm{SEU}_{X(\mathrm{A})} - \mathrm{SEU}_{\bar{X}\,\max(\mathrm{A})}) + (\mathrm{SEU}_{X(\mathrm{S})} - \mathrm{SEU}_{\bar{X}\,\max(\mathrm{S})}) > 0. \qquad [7]$$

It can be seen that the first term in parentheses represents the effects of A on the outcomes while the second term in parentheses represents the effects on outcomes stemming from other factors such as other actors, the impersonal environment, intrinsic satisfactions, etc. The combination of the two then determines whether or not B will perform $X$. Therefore, with respect to B's choice between $X$ and the alternative with the maximum SEU, the first term in parentheses in Expression 7 represents A's realized power over B while the second terms in parentheses represents the effects of situational factors. The combination of the two terms then determines A's effective power over B.

According to this conception of social power, A can exert power over B and have some effect on B's decision and still not obtain B's compliance, for B's ultimate decision and behavior is determined by situational factors, in addition to A's realized power. We could, of course, focus solely on A's effective power and measure power only in terms of B's behavior. However, as Nagel (1968) pointed out, this can lead to confusion for the statement "A has no power to affect B's behavior" can be erroneously interpreted as meaning A is totally impotent and exerts no force at all.

Implicit in our analysis is the idea that power refers to, in Bertrand Russell's words, "the production of intended effects," for the analysis is centered around the behavior that A desires or intends to affect. The central question in our analysis is whether or not A is able to affect B's decision making regarding some behavior A desires. Thus, the assessment of power here requires the determination of *both* A's intentions and the effects A has on B's decision making.

How one defines a concept depends to some extent on the kinds of questions one intends to ask. We feel, however, that the analysis of a number of important issues regarding the use of social power requires some consideration of an actor's intentions and effectiveness in obtaining goals. For example, leadership can be viewed in terms of the exercise of social power (Fiedler, 1968), and questions regarding leader effectiveness require an assessment of how successful the leader is in obtaining goals. In addition various conflicts over the distribution of power in groups and organizations reflect the actors' concerns regarding their ability to obtain their goals. Consequently, we feel that our conception of power may be useful for the analysis of a number of issues. Additional discussion of the issue of intentionality can be found in Cartwright (1965, p. 11) and Wrong (1968, pp. 676–677).

We should point out here that it is not necessary for A to act for B to be aware of A's intentions. B can anticipate A's decisions and intentions on the basis of general cultural knowledge, experiences with others similar to A, communication from other actors regarding A, past experiences with A, etc. Friedrich (1941) referred to this process as the "rule of anticipated reactions," and in his discussion of power in organizations, Simon (1957) wrote:

> an organization member is seldom presented with an ultimatum "to do so and so or suffer the consequences." Rather, he anticipates the consequences of continual insubordination or failure to please the person or persons who have the ability to apply sanctions to him, and this anticipation acts as a constant motivation without expressed threats from any person [p. 196].

Nagel (1968) has pointed out that this "rule of anticipated reactions" is a problem for a number of theories of power, for power is often defined in terms of B's response to some act of A. Consequently, situations in which A does not act fall outside these definitions of power. The effects of anticipated reactions are included in our analysis, for we do not require that B's consideration of A's intentions necessarily results from some specific influence attempt or command by A.

We feel that a consideration of the effects of anticipated reactions in the use of social power helps to clarify the distinction between actual and potential power. Specifically, we argue that the distinction between actual power and potential power can be seen in terms of the distinction between A's power as stemming from an act or influence attempt and A's power as stemming from the effects of anticipated reactions. Thus, both actual and potential power are included in our concept of realized power. Possible power is distinguished from potential power by the fact that possible power refers to the effects A *could have* on B's decision while potential power refers to the effects A *has* through B's anticipation of A's reactions. Wrong (1968) provided further discussion of these distinctions.

Finally, we would like to point out briefly a number of ways in which this analysis may be extended to deal with other issues in the power literature. First, the second term in Expression 7, which represents the effects of situational factors, can be broken down to give separate terms for the realized power of specific other actors. Thus, the analysis may be applied to situations where a number of actors are trying to influence B. Second, as we indicated earlier, the basic ideas of this approach are related to the basic ideas in a number of theories of attitudes and attitude change (Fishbein, 1963, 1967; Peak, 1955; Rosenberg, 1956); consequently, it should be possible to tie in notions of influence through attitude change and manipulation of information with this analysis. Nagel (1968) made the same point. Third, we have focused on B's decision-making process in the most detail since this corresponds with the major focus of the other theories of social power. However, A's and B's decisions are interrelated in that each of the actors is influenced in making his decisions by his perceptions and expectations regarding the other's decision-making process and behavior. A more detailed analysis of this interaction between the decisions of A and B may be useful in clarifying various topics such as bargaining and the effects of counterpower.

## COMPARISON OF VARIOUS THEORIES OF SOCIAL POWER

In this section, the theories of social power discussed above are examined in terms of distinctions that have been made among different aspects and types of social power. This may help to clarify the relationships among these theories by indicating the ways in which the various conceptions of power are similar or different. Our purpose here is not to evaluate theories on the basis of whether or not they include or focus on certain aspects or types of power, but rather to provide a framework for examining the correspondence among the various theories. In some cases, statements made by the various theorists are open to a number of different interpretations. In these situations, we have tried to choose interpretations that are most consistent with the rest of their arguments.

First, it can be seen that the focus in a number of theories is on possible power or the effects A could have on B, while the focus in others is on the effects A actually has on B—realized power or effective power. Lewin, Cartwright, Thibaut and Kelley, and Emerson, along with Dahl, focus on the effects A could have on B. Thus, in terms of their theories, the statement "A has power over B" does not mean that A is necessarily having an effect on B at that time. March, in contrast, is concerned with the effect A actually has on B.

Second, there are differences among the various theories in the extent to which the effects of situational factors are included in the definition of power. We distinguished between A's realized power, which is a function of A's ability to affect outcomes for B, and A's effective power, which is a function of A's realized power *plus* the effects on outcomes stemming from situational factors such as other actors, B's intrinsic reactions to the act, and the impersonal environment. Although some theorists deal only with possible power, we can still look at whether the effects of the use of some of this possible power would be conceptualized in realized or effective terms. Both Lewin's and Cartwright's definitions of power include what we call realized power plus B's negative reactions to being influenced. Cartwright's concept of control, however, includes the effects of various situational factors and is equivalent to our concept of effective power. Thibaut and Kelley's definition of power includes outcomes affected by A and by the situational factors we discussed, and therefore corresponds to our concept of effective power. Emerson's conception of power includes A's realized power *plus* the effects of B's having alternative sources of utilities other than A. March's definition of power (influence) may refer to either realized or effective power since he is concerned with both changes in internal cognitive states and changes in overt behavior. Dahl's definition of power refers to effective power for he is interested in A's ability to affect B's behavior.

Third, there are differences among the various theories of power according to whether or not the definitions of power include the effects of anticipated reactions on the part of B. Cartwright and Dahl defined power as a response to an act of A and thus focus on actual power. Lewin, Thibaut and Kelley, Emerson, and March do not specify that A must act, and consequently the effects of anticipated reactions or potential power can be included in their definitions of power.

## CONCLUSION

Social power has been examined in the context of decision making by the actors involved. We briefly summarize here what we see as the major advantages of this type of analysis.

1. By looking at social power in terms of the perceptions and decisions of the actors, we should be able to make specific, testable predictions about the behavior of actors in power relations. Although we have not dealt with methodological and measurement issues in this paper, successful research in decision theory and related areas suggests that the variables we are dealing with can be fruitfully used in empirical research.
2. The focus on decision making has led us to a consideration of certain factors often overlooked in the analysis of social power, such as subjective probabilities and situational factors, and has suggested ways of incorporating these factors as central aspects of a theory of social power.
3. This approach provides a framework for comparing and perhaps integrating different conceptions of social power. We argued that a number of differences in the various theories of social power correspond to distinctions we made in our analysis.
4. This approach is related to theories in a number of areas of psychology. Consequently, it may be possible to tie in related material on learning, motivation, attitudes, etc., with the analysis of social power.

## REFERENCES

ATKINSON, J. W. *An introduction to motivation.* Princeton: Van Nostrand, 1964.

BACHRACH, P., & BARATZ, M. S. Decisions and nondecisions: An analytical framework. *American Political Science Review*, 1963, **57**, 632–642.

BARCLAY, S. Expected value in psychological models. Department of Psychology Report 7-1-11, University of Washington, Seattle, 1971.

BECKER, G. M., & MCCLINTOCK, C. G. Value: Behavioral decision theory. *Annual Review of Psychology*, 1967, **18**, 239–286.

BLAU, P. M. *Exchange and power in social life.* New York: Wiley, 1964.

CARTWRIGHT, D. Power: A neglected variable in social psychology. In D. Cartwright (Ed.), *Studies in social power.* Ann Arbor, Mich.: Institute for Social Research, 1959. (a)

CARTWRIGHT, D. A field theoretical conception of power. In D. Cartwright (Ed.), *Studies in social power.* Ann Arbor, Mich.: Institute for Social Research, 1959. (b)

CARTWRIGHT, D. Influence, leadership, control. In J. G. March (Ed.), *Handbook of organizations.* Chicago: Rand McNally, 1965.

CARTWRIGHT, D., & ZANDER, A. (Eds.). *Group dynamics* New York: Harper & Row, 1968.

CLARK, K. B. Problems of power and social change: Toward a relevant social psychology. *Journal of Social Issues*, 1965, **21**, 4–20.

COLLINS, B. E., & RAVEN, B. H. Group structure: Attraction, coalitions, communication, and power. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology*. Vol 4. Reading, Mass.: Addison-Wesley, 1969.

DAHL, R. A. The concept of power. *Behavioral Science*, 1957, **2**, 201–218.

DAHLSTRÖM, E. Exchange, influence and power. *Acta Sociologica*, 1966, **9**, 237–284

DULANY, D. E. Hypotheses and habits in verbal "operant conditioning." *Journal of Abnormal and Social Psychology*, 1961, **63**, 251–263.

DULANY, D. E. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. In C. Eriksen (Ed.), *Behavior and awareness*. Durham, N. C.: Duke University Press, 1962.

DULANY, D. E. Awareness, rules, and propositional control: A confrontation with S-R behavior theory. In D. Horton & T. Dixon (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

EDWARDS, W. The theory of decision making. *Psychological Bulletin*, 1954, **51**, 380–417. (Reprinted: In W. Edwards & A. Tversky, *Decision making*. Baltimore: Penguin Books, 1967)

EDWARDS, W. Behavioral decision theory. *Annual Review of Psychology*, 1961, **12**, 473–498. (Reprinted: In W. Edwards & A. Tversky, *Decision making*. Baltimore: Penguin Books, 1967)

EMERSON, R. M. Power-dependence relations. *American Sociological Review*, 1962, **27**, 31–40. (Reprinted: In C. W. Backman & P. F. Secord, *Problems in social psychology*. New York: McGraw-Hill, 1966)

EMERSON, R. M. Operant psychology and exchange theory. In R. L. Burgess & D. Bushell, Jr., (Eds.), *Behavioral sociology*. New York: Columbia University Press, 1969.

EMERSON, R. M. Exchange theory, part one: A psychological basis for social exchange. In J. Berger, M. Zelditch, & B. Anderson (Eds.), *Sociological theories in progress*. Vol. II. Boston: Houghton-Mifflin, in press. (a)

EMERSON, R. M. Exchange theory, part two: Exchange relations, exchange networks and groups as exchange systems. In J. Berger, M. Zelditch, & B. Anderson (Eds.), *Sociological theories in progress*. Vol. II. Boston: Houghton-Mifflin, in press. (b)

FEATHER, N. T. Subjective probability and decision under uncertainty. *Psychological Review*, 1959, **66**, 150–164.

FIEDLER, F. E. Personality and situational deteriminants of leadership effectiveness. In D. Cartwright & A. Zander (Eds)., *Group dynamics*. New York: Harper & Row, 1968.

FISHBEIN, M. An investigation of the relationship between beliefs about an object and the attitude toward that object. *Human Relations*, 1963, **16**, 233–240.

FISHBEIN, M. Attitude and the prediction of behavior. In M. Fishbein (Ed.), *Readings in attitude theory and measurement*. New York: Wiley, 1967.

FRENCH, J. R. P., JR., MORRISON, H. W., & LEVINGER, G. Coercive power and forces affecting conformity. *Journal of Abnormal and Social Psychology*, 1960, **61**, 93–101.

FRENCH, J. R. P., JR., & RAVEN, B. The bases of social power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Institute for Social Research, 1959.

FRIEDRICH, C. J. *Constitutional government and democracy*. Boston: Litte, Brown, 1941.

GAILBRAITH, J., & CUMMINGS, L. L. An empirical investigation of the motivational determinants of task performance. Interactive effects between instrumentality-valence and motivation-ability. *Organizational Behavior and Human Performance*, 1967, **2**, 237–257.

GAMSON, W. A. Experimental studies in coalition formation. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 1, New York: Academic Press. 1964.

GEORGOPOULOS, B. S., MAHONEY, G. M., & JONES, N. W., JR. A path-goal approach to productivity. *Journal of Applied Psychology*, 1957, **41**, 345–353.

GOLDHAMMER, H., & SHILS, E. A. Types of power and status. *American Journal of Sociology*, 1939, **45**, 171–182.

GRAEN, G. Instrumentality theory of work motivation: Some experimental results and suggested modifications. *Journal of Applied Psychology Monograph*, 1969, **53**, (Whole No. 2, Part 2).

HACKMAN, J. R., & PORTER, L. W. Expectancy theory predictions of work effectiveness. *Organizational Behavior and Human Performance*, 1968, **3**, 417–426.

HARSANYI, J. C. Measurement of social power, opportunity costs, and the theory of two-person bargaining games. *Behavioral Science*, 1962, **7**, 67–80. (a)

HARSANYI, J. C. Measurement of social power in *n*-person reciprocal power situations. *Behavioral Science*, 1962, **7**, 81–91. (b)

HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

HOMANS, G. C. The sociological relevance of behaviorism. In R. L. Burgess & D. Bushell, Jr., (Eds.), *Behavioral sociology*. New York: Columbia University Press, 1969.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

HULL, C. L. *A behavior system: An introduction to behavior theory concerning the individual organism*. New Haven: Yale University Press, 1952.

KARLSSON, G. Some aspects of power in small groups. In J. H. Craswell, H. Solomon, & P. Suppes (Eds.), *Mathematical methods in small group processes*. Stanford, Calif.: Stanford University Press, 1962.

KORNBERG, A., & PERRY, S. D. Conceptual models of power and their applicability to empirical research in politics. *Political Science*, 1966, **18**, 52–70.

KUHN, A. *The study of society: A unified approach*. Homewood, Ill.: Irwin, 1963.

LAWLER, E. E. A correlational-causal analysis of the relationship between expectancy attitudes and job performance. *Journal of Applied Psychology*, 1968, **52**, 462, 468.

LAWLER, E. E. *Pay and organizational effectiveness: A psychological view.* New York: McGraw-Hill, 1971.

LAWLER, E. E., & PORTER, L. W. Antecedent attitudes of effective managerial performance. *Organizational Behavior and Human Performance,* 1967, **2**, 122–142.

LEWIN, K. *Field theory in social science.* New York: Harper, 1951.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Roland Press, 1965.

MARCH, J. G. An introduction to the theory and measurement of influence. *American Political Science Review,* 1955, **49**, 431–451.

MITCHELL, T. R., & BIGLAN, A. Instrumentality theories: Current uses in psychology. *Psychological Bulletin,* 1971, **76**, 432–454.

NAGEL, J. H. Some questions about the concept of power. *Behavioral Science,* 1968, **13**, 129–137.

PEAK, H. Attitude and motivation. *Nebraska Symposium on Motivation,* 1955, **3**, 149–159.

PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin,* 1967, **68**, 29–46.

RIKER, W. H. Some ambiguities in the notion of power. *American Political Science Review,* 1964, **58**, 341–349.

ROSENBERG, M. J. Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology,* 1956, **53**, 367–372.

ROTTER, J. B. *Social learning and clinical psychology.* Englewood-Cliffs, N. J.: Prentice-Hall, 1954.

RUSSELL, B. *Power.* New York: Barnes & Noble, 1962.

SCHEIBE, K. E. *Beliefs and values.* New York: Holt, Rinehart & Winston, 1970.

SCHOPLER, J. Social power. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 2. New York: Academic Press, 1965.

SHAPLEY, L. S., & SHUBIK, M. A method for evaluating the distribution of power in a committee system. *American Political Science Review,* 1954, **48**, 787–792. (Reprinted: in M. Shubik (Ed.), *Game theory and related approaches to social behavior.* New York: Wiley, 1965)

SHAW, M. E., & COSTANZO, P. R. *Theories of social psychology.* New York: McGraw-Hill, 1970.

SIMON, H. A. *Models of man.* New York: Wiley, 1957.

TAYLOR, D. W. Decision making and problem solving. In J. G. March (Ed.), *Handbook of organizations.* Chicago: Rand McNally, 1965.

THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups.* New York: Wiley, 1959.

TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century-Crofts, 1932.

TOLMAN, E. C. A cognition motivation model. *Psychological Review,* 1952, **59**, 389–400.

VAN DOORN, J. A. A. Sociology and the problem of power. *Sociologica Neerlandica,* 1962–63, **1**, 3–51.

VROOM, V. H. *Work and motivation.* New York: Wiley, 1964.

WRONG, D. H. Some problems in defining social power. *American Journal of Sociology,* 1968, **73**, 673–681.

ZIPF, S. G. Resistance and conformity under reward and punishment. *Journal of Abnormal and Social Psychology,* 1960, **61**, 102–109.

(Received June 26, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 6, 447–456

# EFFICACY OF PSYCHIATRIC REHABILITATION

WILLIAM A. ANTHONY[1]

*Boston University*

GREGORY J. BUELL, SARA SHARRATT, AND MICHAEL E. ALTHOFF

*Southern Illinois University*

The comparative effectiveness of various psychiatric rehabilitation procedures on posthospital adjustment was surveyed using the outcome criteria of hospital recidivism and posthospital employment. The percentage of psychiatric patients who receive the traditional hospital regimen of drug treatment, and perhaps some form of individual or group therapy, and who are able to remain out of the hospital or find employment was established as a base line against which the unique effects of rehabilitation procedures could be evaluated. Several of the conclusions that evolved out of the survey are as follows: (*a*) Most all types of inpatient treatment innovations improve the patients' in-hospital behavior, but the research does not indicate that these approaches can singularly effect posthospital adjustment. (*b*) Ex-patients who attend aftercare clinics have a lower rate of recidivism than nonattenders. (*c*) Transitional facilities reduce recidivism but demonstrate little effect on enabling patients to function independently, as measured by posthospital employment, in the community. (*d*) There is a definite need for the continued use of specific outcome criteria so that the comparative effectiveness of various psychiatric rehabilitation procedures can be meaningfully evaluated.

The present trend of discharging more and more hospitalized psychiatric patients has prompted a concern about how to best prepare these patients to function adequately when they return to the community (Mosher, Feinsilver, Katz, & Weinckowski, 1970). Reflecting this concern, various rehabilitation procedures have been designed to facilitate the reintegration of the psychiatric patient into the community. The present survey attempts to determine the efficacy of these procedures.

In order to evaluate our current rehabilitation efforts, meaningful outcome criteria must be selected. Typically, procedures designed to rehabilitate the psychiatric patient have been evaluated by using a variety of criteria, including recidivism, posthospital employment, hospital discharge rate, and hospital adjustment. Because of the multiplicity of criteria presently used, it is extremely difficult to compare the effectiveness of different programs.

To help overcome this difficulty, both for this present survey and hopefully to serve as a basis for future comparisons, two criteria, recidivism and posthospital employment, were selected and used consistently throughout the survey. Recidivism is defined as the percentage of discharged psychiatric patients who are subsequently rehospitalized. Posthospital employment, as defined in this survey, refers to the percentage of discharged psychiatric patients who are competitively employed at follow-up, or employed full-time during the entire follow-up period.[2]

In selecting these two criteria, other possible criteria such as hospital discharge rate and various measures of hospital adjustment were ruled out. The research seems to indicate that the evaluations of rehabilitation efforts made from a hospital base do not appear to be consistently related to functioning in the community and thus are not valid outcome criteria if the goals of rehabilitation are to improve community functioning (Bloom &

---

[1] Requests for reprints should be sent to William A. Anthony, Department of Counselor Education, Boston University, 765 Commonwealth Avenue, Boston, Massachusetts 02215.

[2] It should be pointed out that recidivism and posthospital employment are essentially two different criteria. Rehabilitation programs that attempt to have an effect on posthospital employment cannot assume that they will also have an affect on recidivism (Fairweather, 1964; Freeman & Simmons, 1963; Paul, 1969).

Lang, 1970; Ellsworth, Foster, Childers, Arthur, & Kroeker, 1968; Forsyth & Fairweather, 1961; Paul, 1969; Schaeffer, 1969; Walker & McCourt, 1965; Williams & Walker, 1961).

On the other hand, recidivism and posthospital employment seem to be criteria that are more commonly used and more easily defined. This does not imply that there are no problems involved in using these criteria as indicators of community functioning (see Ellsworth et al., 1968). However, the selection of recidivism and posthospital employment as the basic criteria used here does seem to make sense both economically and psychologically. Concerning recidivism, most would agree that one is psychologically "better off" if he can avoid returning to a psychiatric institution; economically, it is less expensive for the taxpayer if the patient remains out of the hospital. Concerning employment, many would agree that working has psychological benefits; economically, work produces tax dollars and avoids government aid payments.

## BASE RATE

In order to evaluate the unique effects of rehabilitation procedures, the effects of traditional hospital treatment on recidivism and posthospital employment must be established as a base line. That is, what is the percentage of psychiatric patients receiving the traditional hospital regimen of drug treatment and perhaps some form of individual or group psychotherapy who are able to remain out of the hospital or find employment. Knowing this base rate, the efficacy of specific rehabilitation procedures can be evaluated by comparing the results obtained by these specific rehabilitation procedures with the results obtained by traditional hospital treatment.

### *Recidivism*

Several studies provide estimates of the recidivism base rate for hospitalized psychiatric patients. Though the studies differ in years sampled, geographic location, and type of institution, their results are remarkably similar, and suggest a recidivism rate for a 1-year period of approximately 40%–50%.

For example, Miller (1966, 1967) and Savino and Schlamp (1968) reported recidivism rates in California for 1956, 1963, 1965, and 1966 to be 40%, 40%, 48%, and 53%, respectively. A study of Chicago residents (Orlinsky & D'Elia, 1964) found a recidivism rate of 46% for a group of patients who received traditional care. A study based at a Cleveland hospital reported 50% readmissions in the first year (Friedman, von Mering, & Hinko, 1966), while an extensive study of Massachusetts' hospitals (Freeman & Simmons, 1963) found a 1-year recidivism rate of 38%.

Other studies not specifically concerned with rate of recidivism report data from which a base rate can be computed. Results of Bloom and Lang (1970) suggest a base rate of 42%, Lorei (1967) a base rate of 39%, while Wilder, Levin, and Zwerling (1966) reported a 45% readmission rate between 1 and 2 years after discharge.

Some researchers also have examined recidivism rates of less than or more than 1 year and typically have reported data that suggest a continual increase in recidivism as time from discharge increases. Within 3 months, reported rates of return were 15% (Orlinsky & D'Elia, 1964); within 6 months, 40%, 33%, and 30% (Fairweather et al., 1960; Friedman et al., 1966; Orlinsky & D'Elia, 1964); within 3 years, 65% (Olshansky, 1968); and within 5 years 67%, 70%, and 75% (Friedman et al., 1966; Miller, 1966, 1967).

### *Employment*

The employment data suggest that during the year following discharge, approximately 20%–30% of the ex-patients work either full-time throughout the year or were employed at the 1-year follow-up date. For example, Hall, Smith, and Shimkunas (1966) classified 28% of their sample as "steady workers," that is, "employed at time of follow-up and had been working continuously on a single job without serious absenteeism [p. 37]." Similarly, Lorei (1967) reported that 24% of his sample had worked full-time for at least 1-year. At 1-year follow-up, Freeman and Simmons (1963) found that 60% of the nonrecidivist patients were still working, or roughly 20% of the total sample of discharged patients.

Base-rate figures for time periods of less or more than 1-year are at best tentative. The only long-term study from which base-rate data can be gleaned reported on a 3-year employment follow-up, which found 25% of the initial sample of hospitalized patients living in the community and competitively employed (Walker, Williams, & Kelley, 1960).

Unfortunately, employment data for a time period of 6 months do not seem to be reported in full-time employment figures. For example, Olshansky (1968) stated that of a sample of 3,248 patients released from 12 Veterans Administration hospitals, 33% worked more than 1 month during the 6 months following discharge. Fairweather (1964) reported 38% of his traditionally treated hospital patients were employed "some of the time," while Walker and McCourt (1965) found that 47% of their sample of Veterans Administration patients had experienced "some employment." However, when Walker and McCourt (1965) used full-time employment as their criterion, the percentage originally classified as employed was decreased by about one-half of 23%.

Thus, the combined base-rate percentages for employment would suggest that during the first 6 months after discharge, 30%–50% of the patients become gainfully employed. When employment percentages are based on either employment status at follow-up or full-time employment during the entire follow-up period, the base rate appears to fall between 20%–30%, regardless of the time period sampled.

Table 1 presents a summary of the base-rate percentages for both the employment and recidivism criteria.[3] Using the base-rate percentages for comparative purposes when necessary, the effectiveness of various rehabilitation procedures can now be surveyed. For purposes of discussion, rehabilitation techniques were divided into two main categories depending on when and where the rehabilitation efforts took place—in the hospital prior to discharge, or after discharge on an outpatient basis.

TABLE 1

SUGGESTED BASE-RATE DATA FOR RECIDICISM AND EMPLOYMENT

| Time since discharge | Recidivism (%) | Full-time Employment (%) |
|---|---|---|
| 6 months | 30–40 | 30–50[a] |
| 12 months | 40–50 | 20–30 |
| 3–5 years | 65–75 | 25[b] |

[a] Percentage of those who became gainfully employed. Rate similar to 12-month figures when criterion is entire 6 months of employment.

[b] Based on only one study.

### INPATIENT REHABILITATION

The hospital rehabilitation programs were further divided into four categories: (*a*) traditional hospital treatment, (*b*) work therapy, (*c*) "total push" therapy, and (*d*) nontraditional groups.

#### *Traditional Hospital Treatment*

It does not seem to matter whether hospitalized psychiatric patients receive eclectically oriented group therapy (Haven & Wood, 1970); psychoanalytically oriented individual or group therapy (Walker & Kelley, 1963); individual or group therapy (Fairweather & Simon, 1963; Fairweather et al., 1960); or drugs, shock, individual or group therapy (Freeman & Simmons, 1963). Regardless of the type of traditional therapy patients receive, their recidivism and employment rates are not differentially affected. In the preceding outcome studies of specific treatment modalities, the recidivism rates were similar to the base-rate figures for hospitalized patients, ranging from a low of 38% for 1 year (Freeman & Simmons, 1963) to a high of 73% for 3 years (Walker & Kelley, 1963). Full-time employment percentages ranged from a high of 42% at 6 months[4]

---

[3] The base-rate data were derived from studies on hospitalized psychiatric patients in general (typically schizophrenics). One researcher (Fairweather, 1964; Fairweather & Simon, 1963; Fairweather, Simon, Gebhard, Weingarten, Holland, Sanders, Stone, & Reahl, 1960) has consistently found that acute psychotics fare better in the community then either hospitalized nonpsychotics or chronic psychotics. If treatment programs deliberately select less severe patients, their recidivism and employment rates should be better than the base-line data presented here.

[4] The validity of this somewhat higher than base-rate percentage is suspect. It was the only one of nine variables that differentiated between treatment groups. Furthermore, by 18 months follow-up the difference in employment has disappeared (Fairweather & Simon, 1963).

(Fairweather et al., 1960) to a low of 23% at 3 years (Walker & Kelley, 1963).

### *Work Therapy*

Kunce (1970) surveyed the literature on work therapy and concluded that the research does not support the idea that work therapy can be therapeutic. He estimated that regardless of whether patients receive work therapy, 33% will become employed, a finding consistent with the base-rate data presented here. A recent study (Walker, Winick, Frost, & Lieberman, 1969), which contrasted two types of work therapy, reported that at a 6-month follow-up 36% of both groups "held regular competitive work at some time during the 6 months."

Some researchers (Barbee, Berry, & Micek, 1969) have suggested that work therapy may foster institutional dependency. Their results indicated that patients participating in work therapy remained longer in the hospital than nonparticipants. During the 2-year follow-up, 46% of the work therapy group and 23% of the nonparticipants were readmitted to the same hospital. However, when readmission to all psychiatric facilities was examined, there were no significant differences between groups.

Another study (Johnson & Lee, 1965) investigated the effects of a hospital program in which work therapy, rehabilitation counseling, and work placement were emphasized. While those investigators did not present figures for recidivism and employment per se, they did report no difference from a control group in mean number of weeks out of the hospital nor community adjustment as rated by heads of the household.

Several studies of work therapy (Cheney & Kish, 1970; Hoffman, 1965) have presented follow-up employment figures considerably better than the base rate. However, these investigators included in their figures only those patients who successfully completed their therapy program or were placed after therapy ended. For example, Hoffman (1965) reported 76% of the patients completing the vocational program at a state hospital were gainfully employed, but when the total population entering the program is included in the analysis, the success figure drops to approximately 36%. Perhaps these programs are counting as successes the 30%–50% who would have found some type of employment even without a hospital-based work therapy program.

The evidence to date indicates that work therapy alone does not increase the patients' probability of remaining in the community or obtaining employment. These conclusions are consistent with the data of Walker and McCourt (1965), who found no relationship between work like activity in the hospital and later community employment.

### *Total Push Therapy*

As defined in this article, total push therapy procedures attempt to structure the total hospital environment so that most all of the patients' waking hours are directed at therapeutic ends. Although these procedures (variously described as milieu therapy, attitude therapy, social learning therapy, and token economies) differ in terms of their theoretical base, the techniques used to facilitate change, and the systematicness of their application, they agree on the necessity of therapeutically structuring the patient's total hospital environment.

In terms of this article, these programs are similar in an even more basic way. While all these approaches have been able to demonstrate positive effects on within-hospital behavior, they have as yet not demonstrated their effects on measures of community adjustment, such as rehospitalization rates (Paul, 1969). The outcome studies that have been done have typically confined their analysis to changes in the patients' ward behavior (e.g., Foreyt & Felton, 1970).

The only total push treatment program that seems to present favorable community outcome data describes itself as a "behavior-milieu" therapy program (Heap, Boblitt, Moore, & Hord, 1970). This treatment approach consisted of a combination of behavior therapy, a token economy, patient government, attitude therapy, and, when needed, other adjunctive programs such as vocational training or psychotherapy. After 35 consecutive months of operation, the behavior-milieu therapy ward's recidivism rate was only 14% as compared to 50% for the rest of the hospital.

The authors were careful to point out that it was difficult to differentiate the unique effects of each treatment. In addition, they suggested that the ultimate clinical effects may not be due primarily to the specific techniques but to the fact that these techniques "set in motion more diffuse interpersonal and social processes that may have powerful and pervasive therapeutic effects [Heap et al., 1970, p. 354]."

### *Nontraditional Groups*

Included in this category are all group approaches, exclusive of group therapy, that seek to improve the patient's interpersonal skills. Fairweather (1964) reported slight support for the use of autonomous problem-solving patient groups in combination with a system in which passes and other responsibilities were contingent upon appropriate behavior. Compared to group therapy patients 6 months after discharge, more of the nontraditional group patients has become "employed some of the time" (54% versus 38%) and less tended to be rehospitalized (36% versus 46%).

Proponents of human relations training groups have reported an effect on both inhospital behavior and posthospital employment and recidivism rates (Hanson, Rothaus, Johnson, & Lyle, 1966; Johnson, Hanson, Rothaus, Morton, Lyle, & Moyer, 1965). Compared to a control group of traditional group therapy patients, the training group has more employed (65% versus 50%) and less rehospitalized (16% versus 26%) at 10-month follow-up. Particularly striking is the unusual finding that even the control group patients' community adjustment seems to be better than the base-rate figures. This inconsistency appears to be due to the fact that the data were obtained by a questionnaire to which only 68% of the total sample responded. Because the control group percentages are incongruent with previously reported base-rate figures, the appropriateness of the experimental group percentages certainly becomes questionable.

Several recent studies (Carkhuff, 1969; Pierce & Drasgow, 1969; Vitalo, 1971) have investigated the effectiveness of systematically training psychiatric patients to function in a more interpersonally facilitative manner and have found that the more structured training group approaches were able to positively effect the within-hospital interpersonal behavior of psychiatric patients. In an extension of this work with adult psychiatric patients to an institutionalized juvenile population, the staff rather than the treatment population was initially trained in interpersonal and program development skills (Carkhuff, 1973). The staff then trained the juveniles in more than 80 skill-learning programs. In addition to significant improvement on indexes of the juveniles' physical, emotional, and intellectual functioning, the 1-year recidivism rate was reduced from 46% to 30%.

## OUTPATIENT REHABILITATION

Rather than trying to prepare the patient for community life while he is still hospitalized, a number of programs have been developed that treat the patient after he has been discharged. These programs can be loosely categorized in terms of the amount of community support that they provide the patient.

### *Drug Maintenance*

Perhaps the minimum expression of support a hospital may provide a discharged patient is to give him a prescription that enables him to continue on medication. However, Williams and Walker (1961) found that the "presence or absence of a prescription for tranquilizing drugs and a report by the patient and his relatives that the drugs were taken regularly as prescribed did not differentiate the recidivists from the non recidivists [p. 90]." The 1-year recidivism rates for a psychiatrically similar group of drug-taking and non-drug-taking patients was 43% and 39%, respectively. Another study (Walker et al., 1960) examined the effects of maintenance medication on the community adjustment of primarily acute schizophrenics 3 months after discharge. Recidivism and employment data did not differentiate the group discharged on drugs from a group discharged without medication.

Two additional investigations compared the rehospitalization rates for a group of patients treated before the use of drugs with a group of patients treated with drugs (Ellsworth &

Clayton, 1960; Michtom, Goldberg, Offenkrantz, & Whittier, 1957). Once again, recidivism rates did not differentiate between the two groups.

### *Aftercare Clinics*

Outpatient or aftercare clinics typically provide the patient with a brief interview plus a review or manipulation of his present level of medication. Although the amount of contact varies somewhat, generally attendance is at least once a month.

Research on these clinics has indicated a significant decrease in recidivism rates for those who attend. Within 6 months to 1 year after discharge, reported recidivism rates have been no higher than 26%, and typically less than 20% (Hornstra & McPartland, 1963; Kris & Carmichael, 1956; Mendel & Rapport, 1963; Orlinsky & D'Elia, 1964; Pollack, 1958; Vitale & Steinback, 1965). For 3, 4, and 5 years, reported rates of recidivism were 20% (Mendel & Rapport, 1963) and 37% (Kris, 1963), respectively, well under the 3 years and over base rate of 65%–75%.

Although patients who attended aftercare clinics are less apt to become rehospitalized, it is not clear why this is so. Clinics differ in terms of what they emphasize, such as a brief interview (Hornstra & McPartland, 1963), existential therapy (Mendel & Rapport, 1963), or home visits by a social worker (Pollack, 1958), yet they all report lower recidivism rates. The one consistent type of treatment is medication, but to explain the reduction in recidivism on the basis of drug treatment alone is to overlook the confounding of this interpretation by patient motivation. That is, the patient's motivation to attend the aftercare clinic may be an expression of his desire to become or stay healthy, and this patient characteristic could be more important than the treatment that he receives.

### *Planned Follow-Up Counseling*

This program differs from aftercare clinics in several ways. First, the treatment philosophy centers less around drug manipulation and more around counseling activities. Second, the analysis of results includes all patients discharged from the hospital into the program, and not just patients motivated enough to participate.

Caffey, Galbrecht, and Klett (1971) randomly assigned patients to two types of aftercare—one group merely was referred to an agency after discharge, while the other group received follow-up counseling, including contact with a social worker who counseled the patient and the individuals with whom he was living, monthly contact with the patient's physician, and continued drug treatment. The recidivism rates reported are slightly lower than the base rate, probably because the patients were screened before they were included in the study. No significant difference in 1-year recidivism rates was found, although the rates did favor the follow-up counseling group (23% to 34%).

The only other investigation of professional follow-up counseling did not report the time period studied, so base-rate comparisons cannot be made (Purvis & Miskimins, 1970). Of interest, however, is the authors' examination of the relationship between amount of patient participation in the follow-up program and recidivism. They concluded that the optimum situation for community adjustment "involves a person who is independent enough to accept some risk (i.e., the separation from the hospital) combined with moderate support. A group in the community or minimum counselor contact represents this type of moderate support . . . [p. 381]."

Summarily, these two studies provide only modest support for the effectiveness of follow-up counseling. The relationship between patient participation in follow-up counseling and recidivism, combined with the positive results of aftercare clinics, seem to suggest that programs with moderate types of community intervention are more apt to exhibit lower rates of recidivism. But once again, this may be due to the characteristics of the patient who seeks mild community support rather than the particular support that he receives.

Dramatic evidence for the effectiveness of follow-up counseling is provided by an investigation of follow-up counseling conducted by nonprofessional volunteers (Katkin, Ginsburg, Rifkin, & Scott, 1971). The treatment approach differed from the previous studies in that unlike the professionals, the volunteer counselors often became extremely involved in many aspects of the patient's life.

The 1-year recidivism rate for female schizophrenics receiving follow-up counseling from the volunteers was 11%, compared to 34% for a noncounseled control group. This significant difference contrasts sharply with the previous follow-up counseling studies, which found a relatively minor effect for less intensive types of community support. Taken together, these three studies seem to suggest that in order to have a dramatic effect on recidivism, follow-up counseling must be flexible enough to provide all levels of community intervention. Discharged psychiatric patients vary in terms of their needs, and 'follow-up counseling must be capable of adapting to a wide variety of patient needs.

### *Transitional Facilities*

Falling under this rubric are such programs variously described as family care, halfway houses, sheltered workshops, and day care centers. For some transitional programs, the goal is to enable the patient to become independent of transitional facility itself, while for others an equally important goal is to enable the patient to become an adjusted member of the transitional facility. Based on this difference in goals, transitional facilities judge their effectiveness at two different times, either after the patient has left the transitional facility or while he still remains dependent on the transitional facility.

Several studies (Lamb & Goertzel, 1971; Wilder, Kessel, & Caulfield, 1968) have followed up patients who have "graduated" from halfway houses to independent living. Recidivism rates 6 months after departure from the halfway houses were 40% (Wilder et al., 1968) and at 18 months, 50% (Lamb & Goertzel, 1971), a relapse rate similar to baserate data. In a more positive vein, Wilder et al. (1968) also reported a full-time employment rate at 6-month follow-up of 53%, a figure considerably better than the base rate.

Rehabilitation programs defined as family care traditionally consist of placing a discharged patient in a private home, a facility that seems to be transitional in a dual sense; that is, from 1 to 2 years after discharge, only a minority (20%–30%) of discharged patients still remain in the home; approximately 30%–35% leave to return to the hospital, and about 40% leave to live independently in the community (Cunningham, Botwinik, Polson, & Weickert, 1969; Molholm & Barton, 1941; Ullmann & Berkman, 1959). By combining those who are living independently with those who are still living in the homes, Ullmann and Berkman (1959) interpreted the effect of family care placement positively; that is, they noted that about 65% are able to stay in the community. Other researchers, however, question whether patients placed in homes are truly participating in community affairs (Lamb & Goertzel, 1971).

Those studies that evaluate the patient's functioning while he is still a member of the transitional facility report favorable data for recidivism, but not for community employment. One study (Beard, Pitt, Fisher, & Goertzel, 1963) of a facility that offered day and evening services reported a slight effect over control group figures for 1-year recidivism rates (35% versus 51%), but no impact on the employment criterion (26% versus 34%). Similarly, Vitale and Steinback (1965) reported a 6-month recidivism rate for their day care patients of only 17%, but also found that only 17% of their day care patients had become employed some of the time.

Meltzoff and Blumenthal (1966) reported on the effectiveness of a day treatment center for patients either referred by an outpatient mental hygiene department or on a trial visit from a hospital. At 18-month follow-up, the recidivism rate for the same type of patients who had received only conventional outpatient treatmnet was 64% as compared to 30% for the day treatment center patients. No conclusions could be drawn about the center's effort on the patient's long-term vocational adjustment.

The most extensive study of a transitional facility compared traditional outpatient aftercare with a "community lodge," where patients lived together as a family, engaged in a group occupational endeavor, and were gradually weaned from most all staff contact

(Fairweather, Sanders, Maynard, & Cresssler, 1969). The employment criterion was the "median percentage of time spent in employment," and the results significantly favored the community lodge group. They point out that this result is to be expected because the "experimental plans deliberately created a situation permitting full time employment [p. 207]." However, the patients' employability did not transfer to competitive employment—of the patients who left the lodge "their employment record outside the lodge strongly resembled that of the control group [p. 207]."

Fairweather et al. (1969) did not report recidivism rates; instead they reported a significant difference in favor of the lodge group for the cumulative number of days an ex-patient remained in the community. However, by combining various numbers from different tables it is possible to compute recidivism rates for both the control and lodge groups; they are practically identical—38% for the control group and 40% for the lodge group.

In trying to summarize the effects of transitional facilities, several consistent findings emerge. Transitional facilities reduce recidivism as long as the patients remain members of the facility. Except for one study (Wilder et al., 1968) no research has demonstrated an effect on competitive community employment once the patient tried to function independently of the transitional facility.

Most transitional facilities, whether they are family care homes, halfway houses, day centers, or lodges, appear to be "transitional" in the sense that the patient "transfers" his dependency from one facility (the hospital) to another (the transitional facility). The main benefits to the discharged patient seem to include more freedom and less stigma; for society, the benefits are somewhat more tangible—it is cheaper (Fairweather et al., 1969).

## CONCLUSIONS AND IMPLICATIONS

1. Although the current emphasis, as exemplified by the community mental health center, is toward avoiding psychiatric hospitalization, it appears that unless or until mental hospitals close their doors individuals will continue to be hospitalized and thus will need to be reintegrated back into the community. While concerted efforts to treat all psychiatric problems on an outpatient basis have been able to reduce significantly the number of people hospitalized (Flomenhaft, Kaplan, & Langsley, 1969; Pasamanick, Scarpitti, & Dinitz, 1967), a significant number of individuals eventually do manage to become hospitalized.

2. Traditional methods of treating hospitalized psychiatric patients, including individual therapy, group therapy, work therapy, and drug therapy, do not affect differentially the discharged patients' community functioning as measured by recidivism and posthospital employment.

3. Most all types of inpatient treatment innovations improve the patients' in-hospital behavior. Yet only two programs, extremely comprehensive in nature, have demonstrated an effect on community functioning (Carkhuff, 1973; Heap et al., 1970). The research does not support the belief that any one of the unique approaches to treatment can singularly effect posthospital adjustment.

4. Aftercare clinics and other forms of moderate community support reduce recidivism. Whether this positive effect is due to the medication administered, the other kinds of services offered, or the type of patient who attends, is not yet clear.

5. Various types of transitional facilities are successful in reducing recidivism but have demonstrated little effect on enabling the patient to function independently in the community as measured by posthospital employment.

6. One of the advantages of the establishment of specific base-rate figures is that it can help identify outcome studies that have used a more easily rehabilitated sample or that have committed a methodological error. This possibility should be examined when the control group has a recidivism or posthospital employment rate that is higher than the base-rate figures presented in this survey.

7. The primary usefulness of specific base-rate criteria is that they provide a basis by which to evaluate and compare existing and future studies in hopes that consistent findings

might emerge. The present authors suspect that some of the unique criteria developed by different investigators may be employed post hoc to maximize the treatment effect. For example, in Fairweather's latest studies (Fairweather, 1964; Fairweather et al., 1969) his experimental group in the first study appears to become his control group in the second. Yet he chose to report his employment and recidivism data differently in each study. Thus, 54% of the experimental group in the first study were "employed some of the time," but when essentially the same treatment group became the control group for his second study on transitional facilities the "median percentage of time spent in full-time employment" for the controls was zero.

The point is not that different methods of follow-up should not be tried but that as unique methods of follow-up are used, they should be used in conjunction with the two criteria outlined in this survey. Although the efficacy of rehabilitation procedures has been evaluated by extensive and well-designed studies, until some uniformity is achieved in reporting results the field remains in many ways an interpretive nightmare. Hopefully, when various rehabilitation programs examine their effects in some consistent manner, more meaningful comparisons of effectiveness can be made between various rehabilitation procedures.

## REFERENCES

Barbee, M. S., Berry, K. L., & Micek, L. A. Relationship of work therapy to psychiatric length of stay and readmission. *Journal of Consulting and Clinical Psychology*, 1969, **33**, 735–738.

Beard, J. H., Pitt, R. B., Fisher, S. H., & Goertzel, V. Evaluating the effectiveness of a psychiatric rehabilitation program. *American Journal of Orthopsychiatry*, 1963, **33**, 701–712.

Bloom, L. B., & Lang, M. E. Factors associated with accuracy of prediction of posthospitalization adjustment. *Journal of Abnormal Psychology*, 1970, **76**, 243–249.

Caffey, E. M., Galbrecht, C. R., & Klett, C. H. Brief hospitalization and after-care in the treatment of schizophrenia. *Archives of General Psychiatry*, 1971, **24**(1), 81–86.

Carkhuff, R. R. *Helping and human relations*. Vols. I & II. New York: Holt, Rinehart, & Winston, 1969.

Carkhuff, R. R. *Cry twice!* From custody to treatment—The story of institutional change. New York: Holt Rinehart, & Winston, 1973.

Cheney, M. T., & Kish, B. G. Job development in a VA hospital. *The Vocational Guidance Quarterly*, 1970, **19**, 61–65.

Cunningham, M. R., Botwinik, W. I., Polson, J. A., & Weickert, A. A. Community placement of released mental patients: A five year study. *Social Work*, 1969, **14**(1), 54–62.

Ellsworth, R. B., & Clayton, W. H. The effects of chemotherapy on length of stay and rate of return for psychiatrically hospitalized patients. *Journal of Consulting Psychology*, 1960, **24**, 50–53.

Ellsworth, R. B., Foster, L., Childers, B., Arthur, G., & Kroeker, D. Hospital and community adjustment as perceived by psychiatric patients, their families, and staff. *Journal of Consulting and Clinical Psychology Monograph Supplement*, 1968, **32**(5), Part 2.

Fairweather, G. W. (Ed.) *Social psychology in treating mental illness*. New York: Wiley, 1964.

Fairweather, G. W., Sanders, D. H., Maynard, H., & Cressler, D. L. *Community life for the mentally ill: An alternative to institutional care*. Chicago: Aldine, 1969.

Fairweather, G. W., & Simon, R. A further follow-up comparison of psychotherapeutic programs. *Journal of Consulting Psychology*, 1963, **27**, 186.

Fairweather, G. W., Simon, R., Gebhard, M. E., Weingarten, E., Holland, J. L., Sanders, R., Stone, G. B., & Reahl, J. E. Relative effectiveness of psychotherapeutic programs: A multicriteria comparison of four programs for three different patient groups. *Psychological Monographs*, 1960, **74**, (5, Whole No. 492).

Flomenhaft, D., Kaplan, M. D., & Langley, G. D. Avoiding psychiatric hospitalization. *Social Work*, 1969, **14**(4), 38–44.

Foreyt, J. P., & Felton, G. S. Change in behavior of hospitalized psychiatric patients in a milieu therapy setting. *Psychotherapy: Theory Research and Practice*, 1970, **7**, 139–141.

Freeman, H. E., & Simons, O. G. *The mental patient comes home*. New York: Wiley, 1963.

Friedman, I., von Mering, O., & Hinko, E. N. Intermittent patienthood: The hospital career of today's mental patient. *Archives of General Psychiatry*, 1966, **14**, 386–392.

Hall, J. C., Smith, K., & Shimkunas, A. Employment problems of schizophrenic patients. *American Journal of Psychiatry*, 1966, **123**, 536–540.

Hanson, P. G., Rothaus, P. M., Johnson, D. L., & Lyle, F. A. Autonomous groups in human relations training for psychiatric patients. *Journal of Applied Behavioral Science*, 1966, **2**, 305–324.

Haven, G. A., & Wood, B. S. The effectiveness of clectic group psychotherapy in reducing recidivism in hospitalized patients. *Psychotherapy: Theory, Research and Practice*, 1970, **7**, 153–154.

Heap, F. R., Boblitt, E. W., Moore, H., & Hord, E. J. Behavior milieu therapy with chronic neuropsychiatric patients: *Journal of Abnormal Psychology*, 1970, **76**, 349–354.

Hoffman, J. J. Paid employment as a rehabilitative technique in a state hospital: A demonstration. *Mental Hygiene,* 1965, **49**, 193–207.

Hornstra, R. K., & McPartland, T. S. Aspects of psychiatric aftercare. *International Journal of Social Psychiatry,* 1963, **9**, 135–142.

Johnson, D. L., Hanson, P. G., Rothaus, P., Morton, R. B., Lyle, F. A., & Moyer, R. Follow-up evaluation of human relations training for psychiatric patients. In E. H. Schein & W. G. Bennis (Eds.), *Personal and organizational change through group methods.* New York: Wiley, 1965.

Johnson, R. G., & Lee, H. Rehabilitation of chronic schizophrenics: Major results of a three-year program. *Archives of General Psychiatry,* 1965, **12**, 237–240.

Katkin, S., Ginsburg, M., Rifkin, M. J., & Scott, J. T. Effectiveness of female volunteers in the treatment of outpatients. *Journal of Counseling Psychology,* 1971, **18**, 97–100.

Kris, E. B. Five-year community follow-up of patients discharged from a mental hospital. *Current Therapeutic Research,* 1963, **5**, 451–462.

Kris, E. B., & Carmichael, D. D. Followup study on patients treated with thorazine. *American Journal of Psychiatry,* 1965, **112**, 1022.

Kunce, J. T. Is work therapy really therapeutic? *Rehabilitation Literature,* 1970, **31**, 297–299.

Lamb, H. R., & Goertzel, V. Discharged mental patients—are they really in the community? *Archives of General Psychiatry,* 1971, **24**, 29–34.

Lorei, T. W. Prediction of community stay and employment for released psychiatric patients. *Journal of Consulting Psychology,* 1967, **31**, 349–357.

Meltzoff, J., & Blumenthal, R. L. *The day treatment center: Principles, application, and evaluation.* Springfield, Ill.: Charles C Thomas, 1966.

Mendel, W. M., & Rapport, S. Outpatient treatment for chronic schizophrenic patients. *Archives of General Psychiatry,* 1963, **8**, 190–196.

Michtom, J., Goldberg, N., Offenkrantz, W., & Whittier, J. Readmission rates for state mental hospital patients discharged on maintenance ataractics. *Journal of Nervous and Mental Disease,* 1957, **125**, 478–480.

Miller, D. Alternatives to mental patient rehospitalization. *Community Mental Health Journal,* 1966, **2**, 124–128.

Miller, D. Retrospective analysis of post-hospital mental patients' worlds. *Journal of Health & Social Behavior,* 1967, **8**(2), 136–140.

Molholm, H. B., & Barton, W. E. Family care, a community resource in the rehabilitation of mental patients. *American Journal of Psychiatry,* 1941, **98**, 33–41.

Mosher, L. R., Feinsilver, D., Katz, M. M., & Weinckowski, L. A. *Special report on schizophrenia.* Bethesda, Md.: National Institute of Mental Health, 1970.

Olshansky, S. The vocational rehabilitation of ex-psychiatric patients. *Mental Hygiene,* 1968, **52**, 556–561.

Orlinsky, N., & D'Elia, E. Rehospitalization of the schizophrenic patient. *Archives of General Psychiatry,* 1964, **10**, 47–54.

Pasamanick, B., Scarpitti, F. R., & Dinitz, S. *Schizophrenics in the community.* New York: Appleton-Century-Crofts, 1967.

Paul, L. Chronic mental patient: Current status—future directions. *Psychological Bulletin,* 1969, **71**, 81–94.

Pierce, R. M., & Drasgow, J. Teaching facilitative interpersonal functioning to psychiatric inpatients. *Journal of Counseling Psychology,* 1969, **16**, 295–298.

Pollack, B. The effect of chlorpromazine in reducing the relapse rate in 716 released patients: Study 3. *American Journal of Psychiatry,* 1958, **114**, 749–751.

Purvis, S. A., & Miskimins, R. W. Effects of community follow-up on post-hospital adjustment of psychiatric patients. *Community Mental Health Journal,* 1970, **6**(5), 374–382.

Savino, M. T., & Schlamp, F. T. The use of non-professional rehabilitation aides in decreasing rehospitalization. *Journal of Rehabilitation,* 1968, **34**(3), 28–31.

Ulmann, L. P., & Berkman, V. C. Efficacy of placement of neuropsychiatric patients in family care. *Archives of General Psychiatry,* 1959, **1**, 273–274.

Vitale, J. H., & Steinback, M. The prevention of relapse of chronic mental patients. *International Journal of Social Psychiatry,* 1965, **11**, 85–95.

Vitalo, R. L. Teaching improved interpersonal functioning as a preferred mode of treatment. *Journal of Clinical Psychology,* 1971, **27**, 166–171.

Walker, R. G., & Kelley, F. E. Short-term psychotherapy with schizophrenic patients evaluated over a three-year follow-up period. *Journal of Nervous and Mental Disease,* 1963, **137**, 348–352.

Walker, R., & McCourt, J. Employment experience among 200 schizophrenic patients in hospital and after discharge. *American Journal of Psychiatry,* 1965, **121**, 316–319.

Walker, R. G., Williams, R. G., & Kelley, F. E. An evaluation of maintenance medication in the post-hospital adjustment of 66 schizophrenic patients. *Journal of Clinical and Experimental Psychopathology,* 1960, **21**, 304–308.

Walker, R., Winick, W., Frost, E. S., & Lieberman, J. M. Social restoration of hospitalized psychiatric patients through a program of special employment in industry. *Rehabilitation Literature,* 1969, **30**, 297–303.

Wilder, J., Kessel, M., & Caulfield, S. Follow-up of a "high-expectations" halfway house. *American Journal of Psychiatry,* 1968, **124**, 1085–1091.

Wilder, J. F., Levin, G., & Zwerling, L. A two year follow-up evaluation of acute psychotic patients treated in a day hospital. *American Journal of Psychiatry,* 1966, **122**, 1095–1101.

Williams, R. A., & Walker, R. G. Schizophrenics at time of discharge. *Archives of General Psychiatry,* 1961, **4**, 87–90.

(Received July 6, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 6, 457–474

# AUTOKINETIC ILLUSION:

## A SYSTEMATIC REVIEW OF THEORIES, MEASURES, AND INDEPENDENT VARIABLES

JOHN LEVY[1]

*Monash University, Melbourne, Australia*

This review attempts to show that in studies of the autokinetic illusion, if the interdependence of theories and measurement methods is considered first, a systematic account becomes possible. It recommends measures based on direction reports because they describe a theoretically significant, unambiguous, immediately apparent aspect of the apparent movement. Two new techniques based on this analysis are also described. The views developed are then applied to a review of experimental studies of the autokinetic illusion. Each type of independent variable is separately considered. The status of the three main current theories—outflow, eye movement, and sensory tonic—is related to this research as well as to other studies of the oculomotor system. The most fruitful areas for future research are indicated.

Previous reviews of the autokinetic illusion (AKI) (Adams, 1912; Crone & Verduyn Lunel, 1969; Reinwald, 1952; Royce, Carran, Aftanas, Lehman, & Blumenthal, 1966) do not resolve the AKI's peculiar measurement problems; therefore, this review first analyzes current AKI theories and measures. From the theoretical account it becomes clear that direction reports are the only measures on which theoretically relevant, unambiguous measures can be based. Two new measures, developed by the author (Levy, 1968), which permit a greater degree of precision, also are introduced. Relevant experiments based on various measures, systematically related by the independent variables used, are then compared, and the theoretical significance of the results indicated. Earlier theories are mentioned, and residual theoretical problems are discussed.

### CURRENT THEORIES

### *Small-Eye-Movement Theory*

In the AKI, subjects describe an actually stationary stimulus as moving. In everyday viewing, retinal displacement is the stimulus for seen movement (Aubert, 1886). In the AKI, retinal displacement can only occur if the eye moves, since the stimulus is fixed. Apparent movement could then result from undetected eye movements. Small eye movements, called "fixation nystagmus," are always present (Heckenmueller, 1965). Since the largest of these are drifts and saccades, averaging 4 minutes of arc, and since the threshold for step displacements is also about 4 minutes of arc (Boyce, 1965), these eye movements cause sufficient retinal displacement to appear as movement if undetected. When fixation nystagmus is eliminated in one dimension, that dimension of the AKI is reduced considerably (Matin & MacKinnon, 1964). Furthermore, since it has a consistent directional bias that varies in every eye, it can be related to consistent individual differences in AKI direction reports (Levy, 1968).

The directional bias is caused by all the orbital forces which keep turning the eye toward a more basic equilibrium position. Of these, the most important is stretch tension from the elastic tissues that make up 50% of the eye muscles' mass (Alpern, 1962). In death, when this is the only effective force, the equilibrium position is called the anatomical position of rest (Stutterheim, 1933). In a conscious subject, muscle tension is modified by many sources of tonic innervation. In the dark, no visual stimuli affect tonus, so, for any fixed body position, a unique equilibrium position, called the physiological position of rest, can be conceived. Levy (1969) measured this position by having subjects report the apparent direction of a brief flash, relative to an earlier one from the same location where the subjects tried to maintain fixation. Errors are known to be mainly due to undetected eye movements (Matin, Pearce, Matin, & Kibler, 1966). The physiological position of rest is operationally defined by that position where, when two-alternative forced-choice decisions are demanded, the second flash is assigned to either side with equal probability; meaning that the eye is equally likely to drift in either direction. The position thus measured remains stable for over a week. Furthermore, an AKI stimulus, placed in the same position, is also equally likely to be reported going in either direction when two-alternative forced-choice decisions are demanded (Levy, 1968). Visual stimuli, like previous eye turn, accommodation, and off-foveal patterns, change fixation nystagmus' direction (Gaarder, 1960; Nachimas, 1961), and hence alter the physiological position of rest. These stimuli affect AKI direction similarly, as outlined below.

---

[1] Requests for reprints should be sent to John Levy, 12 Birriga Road, Bellevue Hill, Sydney, N.S.W., Australia, 2023.

Fixation nystagmus has two major components: a fast and a slow phase, which go in roughly opposite directions. The fast movements (microsaccades) regain foveal fixation after it is lost during the slow drift (Nachimas, 1961). If there was no mechanism to take account of eye movements, both these phases would be equally undetected, their opposite directions would have opposite effects that would cancel each other, and no apparent movement would occur. To explain why, in the AKI, the light seems to move away from the physiological position of rest, it need only be assumed that the slower drifts, toward the physiological position of rest, are not as well detected as the microsaccades. Thus the retinal displacements causing the AKI result from the less detectable drifts. These facts and assumptions are involved implicitly in any small-eye-movement theory of the AKI.

### *Outflow Theory*

The outflow theory includes everything that was outlined above supporting a small-eye-movement theory. However, rather than simply assuming differential sensitivity to the two components of fixation nystagmus, it includes a more comprehensive analysis of how we know our eyes are moving. The basic evidence for this theory comes from studies of large eye movements by Merton (1964). He found that during passive or voluntary eye movements, and with mechanical restraint of the eye during attempted voluntary movement, subjects attributed movement to their eye, or to the visual field, in a pattern that indicated that they always assumed their eyes were stationary unless they attempted to move them. Voluntary efforts were always assumed to be successful. Merton's results are consistent with Helmholtz's (1910) view that only signals from the central nervous system to the eye muscles for voluntary movements (outflows) are monitored. This information, anticipating retinal displacement, is subtracted from the subsequent retinal signal. An outflow account of the AKI assumes that the drifts are involuntary and undetected and consequently cause the apparent movement, while the microsaccades, like their larger counterparts, are monitored and do not contribute to the effect.

The theory is more general than the eye-movement theory since it can explain apparent movement, which occurs during voluntary eye turns when there is no consequent retinal displacement, such as with an afterimage or when the eye is held. It can be extended further to account for other phenomena if pursuit movements also are assumed to be outflow monitored. This latter assumption is justified because, as Rashbass (1959) showed, there is always smooth pursuitlike correction of drifts, as well as microsaccades. This mechanism can be demonstrated only indirectly through the phenomenon of barbiturate nystagmus. In ordinary eye-movement records, drifts and their pursuitlike correction have roughly equal speeds and opposite directions. They cancel each other and cannot be seen separately in the record. The recorded drift component of fixation nystagmus is probably actually the result of at least two opposite forces: a true drift and a pursuitlike correction component. Steinman, Cunitz, Timberlake, and Herman (1967) have shown that, under proper instructions, subjects can increase the pursuitlike component to the extent that microsaccades are not necessary to maintain fixation. In these circumstances the normal AKI is still observed (Steinman, personal communication December, 1969). This report reinforces the idea that the pursuitlike component is outflow monitored, since the AKI occurs when the eye movement pattern shows no direction-specific biases. Other illusions of apparent movement—the optogyral (Byford, 1963), the oculogyral (Graybiel & Hupp, 1946), and the oculogravic (Graybiel & Clark, 1965)—are also marked by little correlation between the recorded eye movements and the reported direction of movement, especially in the later stages of these effects.

The great advantage of the outflow theory, over the small-eye-movement theory, is that it can give a general account of apparent movement in all these cases. A pursuit-like, drift-cancelling response, and an assumption that it is outflow monitored (Ludvigh, 1952a, 1952b), usually are involved, although the research supporting the theory only used saccades (Merton, 1964). Gregory (1958) did present a demonstration that tends to support the idea, and a more comprehensive study by Stoper (1967) suggests that monitoring is partial and develops slowly. The extrapolation is doubtful enough to warrant further study. If it is accepted, the AKI may be explained either by undetected, directionally biased drifts causing retinal displacement and apparent movement in the opposite direction to the drift, or by the outflow monitoring of a pursuitlike cancelling response, which the movement signal initiates.

### *Sensory Tonic Theory*

Sensory tonic theory (Wapner & Werner, 1957; Werner & Wapner, 1952) has two main postulates: First, that the organism continuously releases "sensory tonic energy," which causes sensory distortions when muscular activity is restricted; second, that asymmetrical stimulation shifts spatial orientation norms away from the side of muscle tonus increase and toward an external stimulus.

The theory has logical weaknesses. The AKI's apparent displacement can occur only when changes in the amount of asymmetrical stimulation change spatial norms. In any fixed-stimulus situation, spatial norms are stable, but the AKI moves continuously. The theory cannot account for both effects. Neither can it account for any effects that change direction without a corresponding change in the inducing stimulus, for example, the aftereffect of eye turn on the AKI (Gregory & Zangwill, 1963), the oculogyral illusion (Whiteside, Graybiel, and Niven, 1965), and the Aubert-Müller effect (Day & Wade, 1969). Inter- and intramodal sensory variables are not distinguished. Tonus is treated as a physiological variable, but it has never been measured during experimental tests of the theory and actually functions as an intervening variable (Maze, 1954). Sensory and tonic effects both are attributed to the same cause (sensory tonic energy) merely because they sometimes interact (O'Niel, 1958). The theory gives no account of mechanisms.

Sensory tonic theorists (Comalli, 1960) and others (Fried & Lathrop, 1965; Naylor, 1963) have failed to replicate results on which the theory was based. This may be because the original experiments used inadequate psychophysics. Methodology alone could have produced some results (Weintraub, O'Connell, & McHale, 1964). Other experiments, said to support the theory, ignore equally important contrary evidence (Wishner & Shipley, 1954).

Despite these serious weaknesses, the theory is considered here because it is still invoked frequently. It seems to be an adequate account of the AKI because directional effects can be predicted from the postulates.

## DEPENDENT VARIABLES

### *Verbal Reports*

Most naive subjects spontaneously attribute movement to an AKI stimulus (Graybiel & Clark, 1945). Many subjects say that the apparent displacement is much less than expected from the apparent velocity. The light "moves without going anywhere" (Levy, 1968; Royce et al., 1966, Whiteside et al., 1965). This effect is called paradoxical movement. It makes metric measurement impossible.

Size, distance, shape, brightness, color, and number changes also are reported spontaneously (Cautela & Vitro, 1964). No AKI theory is concerned with them. They can be related to other visual phenomena. Similarly, the autokinetic word technique, in which subjects still report movement when their response category is restricted to written symbols, is mainly used as a projective test (Mednick, Harwood, & Werheim, 1957). It has little value as a report of the AKI's direction since responses may be due to the finer vibrations, rather than the overall tendency, of the AKI's movement.

### *Measurement Criteria*

In good psychophysical measures, subjects make reports that are essentially descriptions but that are limited to a few terms describing a stimulus dimension in which systematic changes have been theoretically predicted. The AKI is primarily an illusion of movement, so measures should be and usually are intended to describe that quality. Movement at any moment is described fully by its velocity, a vector combining speed and direction. However, reports of paradoxical movement show that in the AKI, the movement is different from any in classical mechanics, since apparent displacement is not a product of velocity and time. Hence any request to estimate the amount of movement in a fixed time is ambiguous. Any judgment in an ambiguous situation is biased by suggestion and information that changes subjects' assumptions about that situation (Green & Swets, 1966). Many suggestion-sensitive AKI measures are very reliable when the subjects and test situation are held constant (Gilbert, Hahn, Wackwitz, & Martin, 1965). Whenever the test situation changes, response changes may be due to changes in either the AKI or in the subject's response habits and assumptions (Linton, 1954). Only a signal detection method, as described below, can distinguish these two possibilities.

Good AKI measures therefore must be based on unambiguous judgments that describe a theoretically relevant quality of the movement. Reports of momentary apparent velocity or one of its components, speed or direction, fully describe the AKI at that moment and satisfy the above criteria. No valid speed measures are known. All good AKI measures are therefore essentially direction reports.

### *Nonscalar Measures*

*Direction reports.* Theories mostly predict direction effects, so it is not surprising that direction is the most commonly used AKI response. Like "straight ahead," direction can be judged easily, relative to an implicit standard or norm and then named by familiar terms from the clock face, compass, or angle from vertical. Adams (1912) and Pearce (1966) reported a vector method for recording and averaging large numbers of direction responses in clock histograms. AKI direction reports are reliable and consistent for each subject over long periods if the test situation is stable (Peace, 1966; Wieland & Mefferd, 1966). The most common report is of upward movement (Graybiel & Clark, 1945). Direction is affected by postural factors (Whiteside et al., 1965), especially eye turn (Matin, Pearce, & MacKinnon, 1963). Suggestion has surprisingly little effect (Haggard & Babin, 1948), except where posture is uncontrolled (Sherif, 1948).

*Position of random autokinetic movement.* As eye-turn angle increases, so does the positive correlation between reported AKI and eye-turn direction (Adams, 1912; Matin et al., 1963). When the stimulus is straight ahead, responses are reliable and one direction usually predominates (Pearce, 1966). Levy (1968) found that there is a consistent stimulus position, different for every eye, where each response is equiprobable when subjects make two-alternative forced-choice AKI direction reports. He developed a measure of this "position of random autokinetic movement" (PRAKM), and found that vertical and horizontal dimensions are orthogonal and can be measured independently. He presented stimuli at varying positions along one dimension, using a 5-foot radius perimeter with lights at 5-degree intervals. His measure would have been even more precise if the apparatus had permitted closer stimulus placements. Each stimulus was presented and fixated for 20 seconds to allow the effect of that position to develop fully. After offset, subjects reported the final direction of movement by a simple binary forced-choice response. Only a 5-second interstimulus interval was necessary to record the response and select the next stimulus. Since it is impossible to control eye position without a fixation point (Matin et al., 1966), it would have been better if a stimulus was always present and some other signal, such as a bell, initiated the response.

Many psychophysical methods can be used to estimate the mean of a binary quantal response probability curve. Levy (1968) found that for the AKI, the method of constant stimuli was inefficient, inflexible, and sensitive to aftereffect biases. Wetherill's (1966) modification of the staircase technique was more efficient and better adapted to the specific methodological problems. Stimuli were optimally placed to allow for the large individual differences in mean and variance. The standard error of the mean, which is the optimal starting step size, was estimated to be approximately 5 degrees. Two consecutive PRAKM measures were made, using double random staircases: The first started at straight ahead and ran to 6 reversals, and the second started at the mean of the first and had 10 reversals. They took an average of 32.28 judgments (13.6 minutes) to complete.

Levy's (1968) results showed that this PRAKM measure remains stable for over one week. It was also identical to the physiological position of rest (defined above as that position toward which the eye drifts when the subject attempts to maintain fixation in the dark). The measure therefore can be directly related to the outflow and eye-movement theories since it indicates when the orbital forces are in equilibrium, and unconscious drifts during fixation are equally likely to go in any direction. Adams (1912) and Carr (1910) reported similar stable stimulus positions, differing for each eye, where all AKI direction reports are equiprobable. They used the position as a neutral position to which the effects of other variables could be related, but they did not realize either its theoretical significances or potential as an AKI measure.

Aftereffects from the previous eye position alter the AKI's direction (Gregory & Zangwill, 1963) and may affect many AKI measures. Although aftereffect biases have been eliminated by trial and error from the PRAKM technique suggested above, this technique also has been modified so that the aftereffect was maximized for systematic study (Levy, in presss). An inducing-point source stimulus is presented at a predetermined position for some fixed time. This light is alternated with another at PRAKM; both are fixated continuously while on. Direction reports can be elicited at any time in the cycle by a sound signal. Long-inducing figure inspection in each cycle keeps this at present a "topping up" method, since the evidence indicates that even after 5 minutes at PRAKM the responses have not always returned to the random base line but fluctuate cyclically in a dampened sinusoid form.

Like all AKI direction measures, PRAKM does not indicate speed. It is possible that apparent speed increases with angle of eye turn from PRAKM, but this remains to be established.

### *Scalar Measures*

Since the directional effect of any variable is usually the important question, lack of any size or speed measure is seldom a disadvantage. Nevertheless, more complex scalar measures have been used to compare subjects or conditions in terms of the size of the effect. They all involve serious interpretation difficulties because paradoxical movement and other factors make it impossible to separate sensory and response changes, and they are of no use in testing theories.

*Latency*. Most subjects do not recognize or report AKI movement immediately. Response latencies mostly fall between 5 and 30 seconds, but the distribution is skewed positively, and some naive subjects never respond. Often it is assumed implicitly that latency is a correlate of the rate of development and hence the strength of the AKI. But latency is the least reliable AKI measure (Gilbert, 1967). It is shifted easily by both suggestion and other stimuli (Graybiel & Clark, 1945). Experiments demonstrating latency changes must be interpreted cautiously.

If latency is a true characteristic of the AKI, it implies that the illusion does not occur early in a viewing period. This line of reasoning can be called the "latency hypothesis." It is equally possible that response lags may only be determined by the subjects' criterion of what amount of movement is clear enough to report. Latencies need not occur. In a forced-choice situation, subjects find no difficulty making decisions about direction in a fixed viewing time of only 2.4 seconds (Edwards & Crutchfield, 1951). Even subjects who take 10 minutes to decide, when unrestrained, can be forced to respond in 5 seconds without distorting many effects (Levy, 1968). The latency hypothesis can be tested by a signal detection method, which is described below.

Distance moved depends on both speed and time. Hence, it is traditional to measure latency and then to allow subjects a fixed viewing time after latency so that distance estimates are comparable (Farrow, Santos, Haines, & Solley, 1965). This procedure implicitly accepts the latency hypothesis and also assumes that the AKI occurs in an all-or-none manner. Since both assumptions are unlikely, the procedure is unjustified and compounds any prior confounding.

*Distance, visual angle, and speed.* Many measures assume the AKI is fully described by a path in objective space, mainly in the frontoparallel plane. If so, AKI size could be simply equated with the distance moved and AKI speed could be estimated by dividing distance by viewing time. Royce et al.'s (1966) discussion and evaluation of measurement methods are inadequate because they accept this assumption. They argued:

> An ideal record of autokinetic movement would provide valid measurement of the pattern, length, latency, direction, and instantaneous rate of movement. A method for achieving such an ideal record would give stable results, yet be sensitive and flexible [p. 243].

They do not recognize that, because of paradoxical movement, these measures are mostly ambiguous and cannot be made simultaneously.

The methodologically simplest size measure, verbal estimates of distance moved in a set time, accepts this false assumption. It also assumes that subjects can judge accurately how far away the light is and can name distances reliably. Both tasks are impossible in the dark. Any angular displacement can result from any size real movement, depending on its distance. Movement distance would be especially difficult to judge when the AKI changes direction. Sherif and Cantril (1945), in their study of the effect of suggestion, explicitly chose such judgments for their ambiguity. Identical ambiguity and suggestion effects occur with real movement stimuli (Schonbar, 1945) and with speed estimation (Carr, 1910). Psychophysical speed measures are difficult because movement is relative (Wallach, 1959) and because the simultaneous presence of a comparison stimulus reduces the AKI (Graybiel & Clark, 1945).

If subjects estimate angular displacement, the distance ambiguity is controlled. These visual angle judgments are difficult even with full cues. They too are impossible in the AKI situation, because of paradoxical movement (Graybiel & Clark, 1945).

*Path drawings.* Since the AKI may change direction, a record of the full movement path and size seems to be an improvement on simple direction and distance judgments (Royce et al., 1966). Drawing responses, which are most commonly used, have successfully shown how the AKI's direction changes systematically as a function of some stimulus (Gregory & Zangwill, 1963; Sadler, Mefferd, & Wieland, 1966). This use of drawings differs little from direction reports. More commonly, they have been used to indicate AKI size.

Subjects cannot even draw real movement or familiar figures accurately in the dark. Errors of scale and direction increase with time (Reinwald, 1952). Vertical is drawn 123.7% larger than horizontal (Landauer, Rhine, & Rumiz, 1968). Scale is an arbitrary choice determined by past drawing experience, knowledge of the AKI, and available paper size. It is independent of object size or distance (Blumenthal, 1961) but, like judged size, depends on assumed distance. Drawing size is therefore almost independent of the size of the effect.

Many measures can be taken from drawings. Voth (1941) arbitrarily combined several of them into a single index. Since most of these were related to drawing size, it is not surprising that the index is highly correlated with, and adequately replaced by, the single measure—length of line (Reinwald, 1952; Voth & Mayman, 1963).

Since most drawing parameters depend so much on the subjects' response habits, it is not surprising that they are reliable (Gilbert et al., 1965), even after a year (Haines, Santos, & Farrow, 1965). Since subjects need to make so many assumptions one finds, as expected, that suggestion effectively changes drawing size (Royce et al., 1966). A factor analysis of five drawing and latency scores revealed three distinct factors (Gilbert, 1967). Direction would almost certainly have been a fourth factor. Some experimenters suggest that it would help if all characteristics of the pattern could be assessed by a type of multidimensional scale (Santos, Farrow, & Haines, 1965). These extra suggested scalable qualities are not related theoretically to the causes of the AKI, so they cannot be interpreted directly. It is mostly wasted effort to collect such haphazard data. The author is not aware of any study in which it is used after collection.

*Path at a known scale and distance.* If a subject's scaling behavior can be estimated, AKI size reports can be weighted proportionately. Drawing responses can be related to drawings of a known real movement. Trained subjects who can identify real movements can then apply the same responses to the AKI (Conklin, 1956). Such AKI size reports are biased toward reports about previously observed real movement (Blumenthal, 1961). A subject can describe the AKI path from memory using a reference field, such as a screen, at a set distance (Battersby, Kahn, Pollack, & Bender, 1956; Jordan, 1968). If the light stays at the center of such a screen and subjects fixate the apparent start position, not the light, during the illusion, their new direction of regard can be found after any set time (Cormack, 1963). This effect is smoother, because an off-foveal stimulus biases AKI direction (Carr, 1910). When the stimulus is close to the eye and a drawing board in a frontoparallel plane behind it is within reach, subjects can trace the movement by keeping a pencil tip on the board directly "behind" the light (Royce, Staton, & Kinkade, 1962). This method is as inaccurate as other drawing methods (Reinwald, 1952), possibly because of discrepancies between the apparent visual direction and hand position (Royce et al., 1966), as well as paradoxical movement, common to them all.

*Path as a function of time.* The time of each event is often thought to be essential to a complete AKI record (Royce et al., 1966). It can be recorded if subjects press dots onto their drawing at set intervals. Complex verbal descriptions can be tape-recorded then later timed and described on a grid (Miller & Graybiel, 1962). Such temporal information could be used to show velocity changes within a session, although the distance measures which such estimates would use are biased by paradoxical movement. The analysis seldom is made, even when the data are collected, and would not be useful since predictions about speed are seldom made.

*Compounding with real movement.* Some workers have tried to measure the AKI by a compensatory tracking task in which the subject is given control over the light's position and is told to try and hold the light stationary while his responses are recorded (Bridges & Bitterman, 1954; Davis, 1952). Because of paradoxical movement, the subject's control is partial. He can compensate for either displacement or velocity, but not both, so ambiguities about the size of the effect will remain (Conklin, 1955).

Some real movements of the stimulus are large enough to reverse the AKI direction, suggesting that the interaction between the AKI and real movement involves some monotonically proportionate addition. Since simple direction reports are not confounded by paradoxical movement, it is possible to define a "nulling velocity" as that velocity real movement which, when imposed on the AKI stimulus, reverses the reported direction 50% of the time. In the only use of the method, Pearce and Matin (1966) onset the real movement after the subject reported the first direction of movement; that is, they accepted the latency hypothesis that the AKI appears after a set time at full intensity. This was not necessary. The method would work equally well if the stimulus moved from the start and subjects only made one (terminal) report of apparent direction of movement. Nulling velocity would still be that velocity where the subject is equally likely to report either direction (when two-alternative forced-choice direction reports were the responses) and the confounding with latency could be avoided. Pearce and Matin (1966) measured nulling velocity in eight directions at each stimulus position. Unfortunately, they pooled this data, simply reporting that nulling velocity averaged 20-minute arc per second at straight ahead and 40-minute arc per second with 40 degrees of eye turn. Because of the eye-turn effect, it is likely that nulling velocity varies widely with both eye turn and signal direction. There is no suggestion that nulling velocity measures apparent AKI speed (paradoxical movement separates displacement and velocity effects), but the two variables should increase together monotonically. When modified, as suggested above, the method should be used by any investigator who feels it necessary to scale speed as well as to measure direction. It is an effective method using adequate psychophysics.

*Signal detection methods.* It is possible that the apparent AKI velocity could be assessed directly if real and autokinetic movements were compared successively instead of being presented simultaneously. The task can be paralleled with many sensory discrimination tasks for which signal detection theory provides an appropriate theoretical model. This possibility has not been recognized previously. The AKI is always present and can occur either alone (noise), or algebraically added to a real movement (signal plus noise). Discrimination of the two alternatives depends on the relative sizes of signal and noise and on the criterion stimulus size by which the subject decides which of them has occurred. Four stimulus-response outcomes are possible. If the subject's criterion is manipulated by varying the payoffs he receives for each outcome, sensitivity and criterion can be measured independently; this implies "sensory" and "response" factors can be distinguished (Green & Swets, 1966). Levy (1968) conceived, but has not used, this method.

Most signal detection experiments are concerned with the sensitivity of the observer. Signal, noise, and criterion level are all rigidly controlled to this end. Eijkman, Thijssen, and Vendrick (1966) have developed special techniques that involve varying signal strength and response pattern in such a way that the statistical characteristics of biological noise (here the AKI), and its interaction with other stimulus properties, are analyzed. The intensity-frequency distribution can be measured with different signal sizes; hence their normalcy, and the correlation between signal intensity and noise level, can be assessed. Various signal parameters can be used. In the AKI situation, stimulus size, intensity, duration, or position could be studied separately.

Many signal detection paradigms can be applied immediately to the AKI. Various characteristics of the signal can be left unspecified for the subject to identify. Hence, if the time of onset was left unspecified, the latency hypothesis could be tested. It predicts that the AKI is absent early in a viewing interval. Real movement signals presented early in any interval therefore should be detected more readily than those occurring later, at all criterion levels. It is also possible to test the hypothesis that certain variables, including personality and suggestion (Voth & Mayman, 1963), affect the apparent size of the AKI and not just the response that records it. If the AKI (noise) is greater, sensitivity to a given signal should be less at all criterion levels.

### *Necessary Controls*

There are vestibular, nuchal, and ocular reflexes that affect eye movements and AKI direction. It is therefore essential that the stimulus placement and the subject's position and posture be rigidly controlled. Head position at least should be fixed by a biteboard, and body movement minimized. Aftereffects from eye turn occur and may bias results unless eye position is controlled and stimuli are placed and timed appropriately. Suggestion effects also are induced easily (Royce et al., 1966). Therefore, the situation and instructions must be standardized so that all subjects receive the same type and amount of information about the task. Some confusing experimental results, especially those which suggest voluntary control of the AKI direction, may be due to the omission of these controls (Carr, 1910; Graybiel & Clark, 1945).

## INDEPENDENT VARIABLES

This section reviews systematically the main types of independent variables that have been manipulated in studies of the AKI, and relates the results to the measurement problems and theories that already have been discussed.

### *Ocular Variables*

*Large eye movements.* Some early theorists tested whether the AKI resulted from presumably unconscious, "appropriate" eye movements, "whose length approximately equals the extent of the illusion and whose speed and direction bear some constant relation to the velocity and direction response of the illusion [Carr, 1910, p. 44]." Such movements would cause nonfoveal viewing and poor visibility. Many experiments have shown that they do not occur normally (Bourdon, 1902; Gregory, 1959). Voluntary large eye movements decrease the AKI (Graybiel & Clark, 1945), and greater fixation effort increases it (Gould, 1903). When subjects attempt to fixate the position where movement began, these movements occur voluntarily and are then a response to the AKI and do not cause it (Cormack, 1963).

*Small eye movements.* Some studies of small eye movements (Lehman, 1965; Skolnick, 1940) are only speculative. They use eye-movement measures with 1 degree of minimal sensitivity. Barlow (1952), using better measures in an unrelated study, reported casually that he noticed AKI movements during fixation drifts. Crone and Verduyn Lunel (1969) now have confirmed that drifts of about 1 minute of arc are related consistently to the direction of the AKI. If they are eliminated in any dimension, that dimension of the AKI also is minimized (Matin & MacKinnon, 1964). The importance of fixation nystagmus for the small-eye-movement and outflow theories of the AKI has been outlined above.

*Eye movement when the AKI stimulus goes off.* If the fixation point is turned off after subjects report clear AKI movement in one direction, the eyes drift in the opposite direction, confirming that AKI direction is a function of eye turn, relative to the physiological position of rest. Adams (1912) and Whiteside et al. (1965) measured the movement by turning on the room lights and having subjects report their direction of regard. Merton (1961) used a flash at the original fixation point to produce an afterimage whose retinal position was later determined.

*Knowledge of direction of regard.* Knowledge of direction of regard is poor, which suggests that there is no proprioceptive feedback from the eye muscles, and that only voluntary movements are outflow monitored (Merton, 1964). Viewing the AKI changes it (DeSisto & McLaughlin, 1968; Jordan, 1968) by 4 to 16 degrees (Bourdon, 1902; Graybiel & Clark, 1945). In the Roelof's effect, a moving frame changes the apparent direction of regard and induces apparent movement in a centrally fixated spot. The effect can continue to affect normal AKI direction after frame offset (Brosgole, 1967; Brosgole & Cristal, 1967). This interaction should be studied more fully since the effect might be related to slow-eye-turn responses caused by the fixation reflex (Gaarder, 1960).

*Eye turn.* The probability of the AKI going in the direction of eye turn increases as a function of angle (Adams, 1912; Carr, 1910; Matin et al., 1963). Whiteside et al. (1965) stated that the effect takes 60–90 seconds to develop, but Levy (1968) found that 20 seconds was usually sufficient. Latencies decrease with extreme turns where the AKI is smoother and faster (Carr, 1910, Pearce & Matin, 1966). The discussion of the eye-movement and outflow theories above largely revolves around these experimenters' explanations of this effect, and the measure of PRAKM developed by Levy (1968) also relies on it. The sensory tonic theory also seems to account for the effect, since the stimulation is asymmetrical, but it cannot explain easily why subjects respond consistently when the stimulus is straight ahead (Pearce, 1966).

*Aftereffect of eye turn.* After prolonged eye turn, an AKI stimulus straight ahead usually is said to move in the opposite direction to the previous turn (Adams, 1912; Gregory & Zangwill, 1963). For some subjects, various times and angles of eye turn can also cause aftereffects in the same direction (Carr, 1910; Levy, 1968; in press). The aftereffect reverses after a short time (Gregory & Zangwill, 1963). Preliminary aftereffect studies must anticipate their own results in establishing a procedure free of carry-over effects. Those discussed here therefore may be confounded.

Carr (1910), using direction and velocity reports, found highly variable aftereffects that increased then decreased with inducing time. Direction was a function of angle of previous turn, not always opposite to it. Functions varied between subjects. Gregory and Zangwill (1963) confirmed these results. Carr (1910) felt that these irregular effects could only be explained in terms of two opposing processes.

Davis (1952) reported a study that has many procedural errors, including a compensatory tracking measure, an interstimulus interval that was so short that carry-over effects certainly occurred, and large, inhomogeneous variances that precluded statistical analysis. The results suggested that the aftereffect went in the opposite direction to the previous eye turn throughout the 60-second posttest interval and increased with inducing angle and time.

The aftereffect is usually explained in terms of muscle "strains" or "fatigue" (Adams, 1912; Davis, 1952). These terms originally described the inducing stimulus. A stimulus' name alone does not explain its effect. Gregory and Zangwill (1963) pointed out that weakening of the muscles alone would produce an effect in the same direction as eye turn. Since the opposite direction is the most common first response, they suggest that changes in the "monitoring" or "command" systems are more likely. In other muscle systems, aftereffects of prolonged contraction (from tetanus in isolated nerve/muscle preparations, or isometric strain in the arm, parallel to the ground, that is holding a load) are always a continuation of the contraction response, not weakening or "fatigue." If this also occurs in the eye muscles, the first direction of the aftereffect is predictable. Although no mechanisms have been suggested yet, slowly dissipating temporary changes must be occurring somewhere in the oculomotor system. Ludvigh's (1952a, 1952b) and Gregory and Zangwill's (1963) modifications of the outflow theory include possible mechanisms. These simplified models are still incomplete. It is likely that at least the vergence, pursuit, and saccadic response systems will have to be considered separately to explain the reversal. Since little is known presently of the smooth response systems, systematic studies of the aftereffect's angle, development, and dissipation functions would be valuable. The PRAKM measure outlined above now provides an appropriate technique. Preliminary results confirm Gregory's and Zangwill's (1963) report that the effect oscillates periodically. This may in fact continue for over 5 minutes. Furthermore, the effect is not always in the opposite direction to eye turn. Conditions like excessive convergence induce the opposite effect, confirming Carr's (1910) hypothesis that at least two processes are involved (Levy, in press).

*Blinking.* A fixated point appears to move in the direction of eye turn, because of unconscious eye movements, just before the lid closes, during blinking. A direction of regard where blink movements are random can be measured (Carr, 1907, 1910; Ginsborg & Maurice, 1959). This "blink position" is a stable characteristic of each eye. It is identical to PRAKM (Adams, 1912; Carr, 1910), which is itself identical to the physiological position of rest (Levy, 1968).

### *Visual Stimulus Variables*

*Retinal stimulus stationary.* Horizontal AKI movement is considerably reduced when the retinal image is stabilized in that dimension (Matin & MacKinnon, 1964). No other studies of stabilized retinal images (Heckenmueller, 1965) make any mention of apparent movement. Afterimages are also fixed on the retina and move with the eye. During monitored saccadic effort, if the eye is not restrained, apparent movements of an afterimage are associated with appropriate eye movements (Gregory, 1958; Whiteside et al., 1965). Off-foveal afterimages appear to move continuously to the periphery as saccades attempting foveal fixation occur (Rexroad, 1928). No apparent movement occurs during drifts and passive movement. All these findings are explained by the outflow theory, but not the sensory tonic or eye-movement theories.

*AKI relative to a retinally stationary stimulus.* An afterimage, viewed simultaneously with an AKI stimulus, can be used to demonstrate that large eye movements do not occur since the two stimuli do not separate (Marx, 1913). Movement is relative (Wallach, 1959), so subjects attribute the slight displacements that do occur to either or both stimuli. Since the outflow theory accounts for the movements of either stimulus alone, it accounts for their relative movements too. The small-eye-movement theory cannot explain why they move together, and the sensory tonic theory why they move independently.

*Shape.* Symmetrical, centrally fixated, geometrical figures move normally (Adams, 1912), but shape can affect AKI direction (Graybiel & Clark, 1945). Comalli, Werner, and Wapner (1957) attributed shape effects to dynamic properties resulting from meaning. They found direction of movement of ambiguous figures depended on the interpretation suggested (e.g., balloon/parachute), but Comalli (1960) failed to replicate his own result. Adams (1912) got no dynamic effect from lines with arrowheads. Any direction effects may well result from the pattern being asymmetrical about the fovea after suggestion directs attention to different parts.

*Asymmetry about the fovea.* When a figure is fixated at one end or an extra light is visible (Adams, 1912; Sadler et al., 1966), the AKI movement is in the opposite direction to the asymmetry. The effect probably is due to the fixation reflex,[2] which biases the drift component of fixation eye movements toward the center of the figure (Gaarder, 1960). Sensory tonic theory would explain it as a shift of "norms" toward the center of an asymmetrical stimulus.

*Size.* Size has always been confounded with luminous flux in AKI studies because light sources have been constant. Size increases have the same effect as intensity; latency increases and size estimates decrease (Edwards, 1954a, 1954b; Karowski, Redner, & Wood, 1948). The size effect is not linear. Reinwald (1952) found stimuli between .4 and 4.0 millimeters diameter, at an unspecified distance, had no effect. Schilder (1912) and Adams (1912) used stimulus diameters of 1–5 centimeters. Both found a sharp increase occurred between 3 and 4 centimeters; this becomes 1 degree at the 200-centimeter distance used by Adams (1912). Contours of larger stimuli fall into the less sensitive periphery. The AKI reduction with larger stimuli therefore can be explained. Furthermore, it is likely that the fixation reflex (see Footnote 2) increases with large patterned stimuli. Thus stability of the everyday visual field is not a major theoretical problem, as Carr (1910), Gregory and Zangwill (1963), and Howard and Templeton (1966) suggest. Large, patterned stimuli do, in fact, move a little after long fixation. Changes in size, shape, distance, and brightness also are reported (Edwards, 1954a, 1954b; Honisett & Oldfield, 1961).

*Exner's experiment.* Exner (1896) reported that when a black disk with a pinhole in it was dimly illuminated from behind and the hole was fixated, the dot and disk moved independently. Eye-movement AKI theories implied both should always move together, so Exner rejected them and proposed a retinal fatigue theory (Reinwald, 1952), which anticipated satiation theory (Kohler & Wallach, of reference [p. 401]." Others believe that apparent fluctuation of stable sensory input is a central characteristic of certain personality types (Comalli, 1960; Voth, 1962). It is not sufficient to say the AKI occurs because something is absent. These theories do not identify the mechanism or eliciting stimulus and hence cannot explain the AKI.

---

[2] The term "fixation reflex" is used ambiguously (Davson, 1963). It may mean the process by which foveal fixation is achieved by a saccadic movement. It can also mean a process by which the eye moves slowly and unconsciously toward the center of the visual array. The latter sense is intended here. The terms "centering reflex" or, in optokinetic nystagmus, "following reflex" can also be used. In some cases of brain damage, the process is so dominant that fixation cannot be removed (Davson, 1963).

## THEORETICAL IMPLICATIONS

### *Deficiencies of the Sensory Tonic Theory*

Although sensory tonic theory seems to predict successfully the effect of some variables on AKI direction, alternative, more specific explanations are always available. Sensory tonic theory does not really explain the AKI. Since it only predicts changes in spatial norms with changes in muscle tonus, it cannot explain why the AKI's position seems to change continuously and increasingly as subjects become more stationary and their "muscular tonic balance" more fixed. Sometimes successful predictions are due to illegitimate reasoning; for example, Miller, Werner, and Wapner (1958) predicted that as the frequency of a tone went higher or lower, an AKI stimulus, which was present at the same time, would do likewise. The theory actually is concerned only with the effects of stimuli that change spatial location. Often predictions about direction seem arbitrary and specious because tonus is such an imprecisely defined stimulus; for example, Wishner and Shipley (1954) concluded handedness (said to indicate tonic balance) affected AKI direction, though this result occurred irrespective of the hand used to draw the response. This study also ignores equally important contradictory evidence (the effect does not occur for left handers). Several effects cannot be accounted for at all; for example, an afterimage and an AKI stimulus sometimes move independently. Although no current theory can account for the aftereffect of eye turn, it presents special problems for sensory tonic theory since it changes direction without a corresponding change in the stimulus (Gregory & Zangwill, 1963). Because the theory has logical weaknesses and is based on terms that have no real operational correlates, it cannot be related to the oculomotor system. Therefore, it is not an adequate, comprehensive account of the AKI.

### *Deficiencies of the Eye-Movement Theory*

Other explications of the eye-movement theory have not recognized that there were consistent directional biases in fixation nystagmus, let alone explained them. As presented here, the concept of the physiological position of rest, and the changes in this equilibrium position that occur as a result of stimulation, have developed the theory to the point where it explains most AKI research.

Small eye movements normally do occur (Crone & Verduyn Lunel, 1969). They may not always do so, since an afterimage and an AKI stimulus sometimes move together. Steinman's (personal communication December, 1969) report that the AKI occurs even when saccadic correction is eliminated from fixation nystagmus also suggests that the small-eye-movement theory alone is not sufficient to explain all specific AKI effects. Neither can it account for other, undoubtedly related, illusions of movement, such as the apparent movement of an afterimage (Whiteside et al., 1965). Further, Byford (1963) has shown that eye movements and apparent movement direction are not always related in the optogyral illusion, so it is equally unlikely that they will be related in other complex effects, such as the aftereffect of eye turn on the AKI (Levy, in press). The account developed here would predict that eye movements and AKI direction are only likely to be correlated in the early phases of these effects when naive subjects experience them for the first time. In later phases, monitored pursuits would eliminate the correlation from the record, while still producing the illusion.

### *Deficiencies of the Outflow Theory*

The outflow theory is a general theory of oculomotor control, which attempts to explain many cases of apparent movement. In the introduction it was argued that its exposition was usually simple and schematic. Outflow monitoring has only been demonstrated empirically with saccadic movements. A slow, pursuitlike stabilizing response is involved in most cases of apparent movement where the outflow theory has a distinct advantage over a simple eye-movement theory. These slow responses, as yet, are studied little. The implicit assumption that they are also outflow-monitored still requires confirmation. Stoper (1967) presented evidence that at least some pursuit monitoring occurs.

Other types of slow changes in oculomotor control occur that the outflow theory cannot explain: Version size quickly adapts when the normal contingency between eye movements and the consequent retinal displacement is altered (McLaughlin, 1967). Pursuits improve with practice (Fender, 1964). Eventually sinusoidal movements can, for a short time, be produced voluntarily after the usually necessary target is removed (Deckert, 1964). Versions are anticipatory and use more information than is available in the retinal signal one reaction time earlier. By linear extrapolation of target path, a moving target can be fixated with .06-second effective lag (Robinson, 1965). The physiological position of rest is adjusted relatively permanently during behavioral compensation (Day & Singer, 1967) to altered visual-tactual relations such as prism adaptation (McLaughlin & Webster, 1967) and hidden phoria (Alpern, 1962). Although the outflow theory has been modified and expanded (Ludvigh, 1952b; Milhorn, 1966), these accounts are still incomplete and do not consider or explain the above effects. There are more processes involved in the oculomotor system than these models explicitly recognize (Levy, in press).

## Conclusions

When AKI measurement methods are analyzed, it is clear that any current measures, other than those based on direction reports, are ambiguous. Experiments that use ambiguous measures are confounded by response biases. The results of studies using such measures can often be reduced to "suggestion" effects. Most other AKI experiments can be explained equally by the small-eye-movement and outflow theories when some basic assumptions about the physiological position of rest and the effect it has on the direction of fixation nystagmus are explicit. The outflow theory is a more general account of mechanisms in the oculomotor system. If usually implicit assumptions about the nature of slow movements are included, cases where AKI and fixation nystagmus direction are not correlated, and many other illusions of movement, can be explained by the same model. Information about short- and long-term regulation of oculomotor equilibrium will increase our understanding of apparent movement direction, as well as oculomotor control in general.

## REFERENCES

Adams, H. F. Autokinetic sensations. *Psychological Monographs,* 1912, **14**, 1–45.

Alpern, M. Movements of the eyes. In H. Davson (Ed.), *The eye.* Vol. 3. Part 1. New York, Academic Press, 1962.

Aubert, H. Die Bewegungsempfindung. *Archiv für die Gesamte Physiologie,* 1886, **39**, 347–370.

Barlow, H. B. Eye movements during fixation. *Journal of Physiology,* 1952, **116**, 290–306.

Battersby, W. S., Kahn, R. L., Pollack, M., & Bender, M. B. Effect of visual, vestibular, and somatosensorimotor deficit on autokinetic perception. *Journal of Experimental Psychology,* 1956, **52**, 398–410.

Blumenthal, A. L. Measurement of the autokinetic illusion. Unpublished master's thesis, University of Washington, 1961.

Bourdon, B. *La perception visuelle de l'espace.* Paris: Scheicher frères, 1902.

Boyce, P. R. The visual perception of movement in the absence of an external frame of reference. *Optica Acta,* 1965, **12**, 47–54.

Bridges, C. C., & Bitterman, M. E. The measurement of autokinetic movement. *American Journal of Psychology,* 1954, **67**, 525–529.

Brosgole, L. Induced autokinesis. *Perception and Psychophysics,* 1967, **2**, 69–73.

Brosgole, L., & Cristal, R. M. Vertically induced autokinesis. *Psychonomic Science,* 1967, **7**, 337–338.

Byford, G. M. Eye movements and the optogyral illusion. *Aerospace Medicine,* 1963, **34**, 119–123.

Carr, H. A. A visual illusion of motion during eye closure. *Psychological Review Monograph Supplement,* 1906, **7** (3, Whole No. 31)

Carr, H. A. The autokinetic sensation. *Psychological Review,* 1910, **17**, 42–75.

Cautela, J., & Vitro, F. The effect of instructions on the appearance of the autokinetic effect. *Journal of Psychology,* 1964, **58**, 85–88.

Chaplin, J. Sex differences in the perception of autokinetic movement. *Journal of General Psychology,* 1955, **52**, 149–155.

Comalli, P. E., Jr. Studies in physiognomic perception IV. Differential effects of directional dynamics of pictured objects on real and apparent motion in artists and chemists. *Journal of Psychology,* 1960, **49**, 99–109.

Comalli, P. E., Werner, H., & Wapner, S. Studies in physiognomic perception III. Effect of directional dynamics and meaning induced sets on autokinetic motion. *Journal of Psychology*, 1957, **43**, 289–299.

Conklin, J. E. A note on the measurement of autokinetic movement. *American Journal of Psychology*, 1955, **68**, 144–145.

Conklin, J. E. Quantitative determination of autokinetic movement; Rate, direction, extent, and latency. *Northwest Science*, 1956, **30**, 61–69.

Conklin, J. E., The influence of figural inspection on the autokinetic illusion. *American Journal of Psychology*, 1957, **70**, 395–402.

Cormack, R. H. Ocular tracking as a measure of autokinetic movement. *Perceptual and Motor Skills*, 1963, **17**, 223–226.

Crone, R. A., & Verduyn Lunel, H. F. E. Autokinesis and the perception of movement: The physiology of eccentric fixation. *Vision Research*, 1969, **9**, 89–102.

Davis, W. A. A "muscle strain" theory of autokinesis. Unpublished master's thesis, University of Washington, 1952.

Davson, H. *Physiology of the eye*. London: Churchill, 1963.

Day, R. H., & Singer, G. Sensory adaptation and behavioral compensation with spatially transformed vision and hearing. *Psychological Bulletin*, 1967, **67**, 307–322.

Day, R. H., & Wade, N. J. Mechanisms involved in visual orientation constancy. *Psychological Bulletin*, 1969, **71**, 33–42.

Deckert, G. H. Pursuit eye movements in the absence of a moving stimulus. *Science*, 1964, **143**, 1192–1193.

DeSisto, M. J., & McLaughlin, S. C. Changes in judgment of direction of gaze after autokinetic perception. *Psychonomic Science*, 1968, **10**, 221–222.

Edridge-Green, F. W. Visual phenomena connected with the yellow spot. *Journal of Physiology*, 1910, **41**, 263–275.

Edwards, W. Two- and three-dimensional autokinetic movement as a function of size and brightness of stimuli. *Journal of Experimental Psychology*, 1954, **48**, 391–398. (a)

Edwards, W. Autokinetic movement of very large stimuli. *Journal of Experimental Psychology*, 1954, **48**, 493–495. (b)

Edwards, W. Information and autokinetic movement. *Journal of Experimental Psychology*, 1959, **57**, 89–90.

Edwards, W., & Crutchfield, R. S. Differential reduction of autokinetic movement by a fixated figure. *Journal of Experimental Psychology*, 1951, **42**, 25–31.

Eijkman, E., Thijssen, J. M., & Vendrick, A. J. H. Weber's law, power law, and internal noise. *Journal of the Acoustical Society of America*, 1966, **40**, 1164–1173.

Exner, S. Über autokinetische empfindungen. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 1896, **12**, 313–330.

Farrow, B. J., Santos, J. F., Haines, J. R., & Solley, C. M. Influence of repeated experience on latency and extent of autokinetic movement. *Perceptual and Motor Skills*, 1965, **20**, 1113–1120.

Fender, D. H. Control mechanisms of the eye. *Scientific American*, 1964, **211**(1), 24–33.

Feree, C. E. The streaming phenomenon. *American Journal of Psychology*, 1908, **19**, 484–503.

Fried, R., & Lathrop, R. G. Effects of extraneous stimulation on the visual perception of verticality: A failure to replicate. *Journal of Experimental Psychology*, 1965, **69**, 327–328.

Gaarder, K. Relating a component of physiological nystagmus to visual display. *Science*, 1960, **132**, 471.

Gilbert, D. C. A factor-analytic study of autokinetic responses. *Journal of Experimental Psychology*, 1967, **73**, 354–357.

Gilbert, D. C., Hahn, J. P., Wackwitz, J. H., & Martin, R. M. Stability of autokinetic movement. *Perceptual and Motor Skills*, 1965, **21**, 715–718.

Ginsborg, B. L., & Maurice, D. M. Involuntary movements of the eye during fixation and blinking. *British Journal of Opthalmology*, 1959, **43**, 435–437.

Goldman, A. E. Studies in vicariousness: Degree of motor activity and the autokinetic phenomenon. *American Journal of Psychology*, 1953, **66**, 613–617.

Gould, G. M. An hitherto undescribed visual phenomenon. *Science*, 1903, **18**, 536–537.

Graybiel, A., & Clark, B. The autokinetic illusion and its significance in night flying. *Journal of Aviation Medicine*, 1945, **16**, 111–151.

Graybiel, A., & Clark, B. Validity of the oculogravic illusion as a specific indicator of otolith function. *Aerospace Medicine*, 1965, **36**, 1173–1181.

Graybiel, A., & Hupp, D. The oculogyral illusion. *Journal of Aviation Medicine*, 1946, **17**, 3.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

Gregory, R. L. Eye movements and the stability of the visual world. *Nature*, 1958, **182**, 1214–1216.

Gregory, R. L. A blue filter technique for detecting eye movements during the autokinetic effect. *Quarterly Journal of Experimental Psychology*, 1959, **11**, 113–114.

Gregory, R. L., & Zangwill, O. L. The origin of the autokinetic effect. *Quarterly Journal of Experimental Psychology*, 1963, **15**, 252–261.

Guilford, J. P. Autokinesis and the streaming phenomena. *American Journal of Psychology*, 1928, **40**, 401–417.

Guilford, J. P., & Dallenbach, K. M. A study of the autokinetic sensation. *American Journal of Psychology*, 1928, **40**, 83–91.

Haggard, E. A., & Babin, R. On the problem of reinforcement in conditioning the autokinetic phenomenon. *Journal of Experimental Psychology*, 1948, **38**, 511–525.

Haines, J. R., Santos, J. F., & Farrow, B. J. Stability of the autokinetic phenomenon. *Perceptual and Motor Skills,* 1965, **21**, 394.

Heckenmueller, E. G. Stabilization of the retinal image: A review of method, effects, and theory. *Psychological Bulletin,* 1965, **63**, 157–169.

Helmholtz, H. von. Handbuch der Physiologischen optik. (3rd ed.) Leipzig: Voss, 1910 (Republished: J. C. P. Southall (Trans.) *A treatise on physiological optics.* New York: Dover, 1962).

Honisett, J., & Oldfield, R. L. Movement and distortion in visual patterns during prolonged fixation. *Scandinavian Journal of Psychology,* 1961, **2**, 49–55.

Howard, I. P., & Templeton, W. B. *Human spatial orientation.* London: Wiley, 1966.

Hunter, W. S. The after-effect of visual motion. *Psychological Review,* 1914, **21**, 250–253.

Jordan, S. Autokinesis and felt eye position. *American Journal of Psychology,* 1968, **81**, 497–512.

Karowski, T. F., Redner, H., & Wood, H. O. Autokinetic movement of very large stimuli. *Journal of General Psychology,* 1948, **39**, 29–37.

Koffka, K. *Principles of Gestalt psychology.* London: Routledge & Kegan-Paul, 1935.

Kohler, W., & Wallach, H. Figural aftereffects: An investigation of visual processes. *Proceedings of the American Philosophical Association,* 1944, **88**, 269–357.

Kylstra, J. Autokinesis. *Aeromedica Acta,* 1956–1957, **5**, 261–270.

Landauer, A. A., Rhine, C. A., & Rumiz, L. Overestimation of vertical movement. *Psychonomic Science,* 1968, **10**, 59–60.

Lehman, R. S. Eye movements and the autokinetic illusion. *American Journal of Psychology,* 1965, **78**, 490–493.

Levy, J. L. The measurement and explanation of the directional component of the autokinetic illusion. Unpublished master's thesis, Monash University, Australia, 1968.

Levy, J. L. The physiological position of rest and phoria. *The American Journal of Opthalmology,* 1969, **68**, 706–713.

Levy, J. L. Autokinesis direction during and after eye turn. *Perception and Psychophysics,* in press.

Linton, H. B. Autokinetic judgments as a measure of influence. *Journal of Abnormal and Social Psychology,* 1954, **49**, 464–465.

Livson, N. H. Aftereffects of prolonged inspection of apparent movements. *American Journal of Psychology,* 1953, **66**, 365–376.

Luchins, A. S. The autokinetic effect and gradation of illumination of the visual field. *Journal of General Psychology,* 1954, **50**, 29–37.

Ludvigh, E. J. Possible role of proprioception in the extraocular muscles. *Archives of Opthalomology,* 1952, **48**, 436–441. (a)

Ludvigh, E. J. Control of ocular movements and visual interpretations of environment. *Archives of Opthalmology,* 1952, **48**, 442–448. (b)

Marx, E. Unterschungen uber fixation unter verschieden—en bedingungen. *Zeitschrift für Sinnesphysiologie,* 1913, **47**, 79–96.

Matin, L., & MacKinnon, G. E. Autokinetic movement: Selective manipulation of directional components by image stabilization. *Science,* 1964, **143**, 147–148.

Matin, L., Pearce, D., & MacKinnon, G. E. Variation of directional components of autokinetic movement with position of eye in the orbit. *Journal of the Optical Society of America,* 1963, **53**, 521.

Matin, L., Pearce, D., Matin, E., & Kibler, G. Visual perception of direction in the dark. Roles of local sign, eye movement, and ocular proprioception. *Vision Research,* 1966, **6**, 453–469.

Maze, J. R. Do intervening variables intervene? *Psychological Review,* 1954, **61**, 226–234.

McKitrick, K. G. Bodily activity and perceptual activity. *Perceptual and Motor Skills,* 1965, **20**, 1109–1112.

McLaughlin, S. C. Parametric adjustment in saccadic eye movements. *Perception and Psychophysics,* 1967, **2**, 359–362.

McLaughlin, S. C., & Webster, R. C. Changes in the straight ahead position during adaptation to wedge prisms. *Perception and Psychophysics,* 1967, **2**, 37–44.

Mednick, S., Harwood, A., & Werheim, J. Perception of neutral and disturbing words through the autokinetic word technique. *Journal of Abnormal and Social Psychology,* 1957, **55**, 267–268.

Merton, P. A. The accuracy of directing the eyes and the hand in the dark. *Journal of Physiology,* 1961, **156**, 555–557.

Merton, P. A. Absence of conscious position sense in the human eyes. In M. A. Bender (Ed.), *The oculomotor system.* New York: Harper and Row, 1964.

Milhorn, H. T. *The application of control theory to physiological systems.* Philadelphia: Saunders, 1966.

Miller, E. F., & Graybiel, A. Comparison of autokinetic movement perceived by normal persons and deaf subjects with bilateral labyrinthine defects. *Aerospace Medicine,* 1962, **33**, 1077–1080.

Miller, A., Werner, H., & Wapner, S. Studies in physiognomic perception V: Effect of ascending and descending gliding tones on autokinetic motion. *Journal of Psychology,* 1958, **46**, 101–105.

Nachimas, J. Determiners of the drift of the eye during monocular fixation. *Journal of the Optical Society of America,* 1961, **51**, 761–766.

Naylor, G. F. K. Effects of stress on the perception of direction. *American Journal of Psychology,* 1963, **15**, 17–28.

O'Niel, W. M. Basic issues in perceptual theory. *Psychological Review,* 1958, **65**, 348–361.

Pearce, D. G. Consistency of individual patterns of autokinetic direction. *Perceptual and Motor Skills,* 1966, **23**, 1119–1123.

Pearce, D. G., & Matin, L. The measurement of autokinetic speed. *Canadian Journal of Psychology,* 1966, **20**, 160–172.

Rashbass, C. Barbiturate nystagmus and the mechanisms of visual fixation. *Nature,* 1959, **183**, 897.

Reinwald, F. L. An experimental investigation of conditions affecting autokinetic sensations. Unpublished doctoral dissertation, University of Texas, 1952.

Rexroad, C. N. Eye movements and visual afterimages. *American Journal of Psychology,* 1928, **40**, 426–433.

Robinson, D. A. The mechanisms of human smooth pursuit eye movements. *Journal of Physiology,* 1965, **180**, 569–591.

Royce, J. R., Carran, A. B., Aftanas, M., Lehman, R. S., & Blumenthal, A. The autokinetic phenomenon: A critical review. *Psychological Bulletin,* 1966, **65**, 243–260.

Royce, J. R., Staton, W. R., & Kinkade, R. G. Experimental reduction of autokinetic movement. *American Journal of Psychology,* 1962, **75**, 221–231.

Sadler, T. G., Mefferd, R. B., & Wieland, B. A. Extent, direction, and latency of autokinetic movement as a function of placement of an adjacent light. *Perceptual and Motor Skills,* 1966, **23**, 1087–1096.

Santos, J. F., Farrow, B. J., & Haines, J. R. Consistency and style of autokinetic movement. *Perceptual and Motor Skills,* 1965, **21**, 583–586.

Schilder, P. Über autokinetsche empfindung. *Archiv für Gesamte Psychologie,* 1912, **25**, 36–77.

Schonbar, R. A. The interaction of observer pairs in judging visual extent and movement. The formation of social norms in structured situations. *Archives of Psychology,* 1945, **41**, 299.

Sexton, M. C. The autokinetic test: Its value in psychiatric diagnosis and prognosis. *American Journal of Psychiatry,* 1945, **102**, 399–402.

Sherif, M. *An outline of social psychology.* New York: Harper, 1948.

Sherif, M., & Cantril, H. The psychology of attitudes. Part 1. *Psychological Review,* 1945, **52**, 293–319.

Singh, S. D., & Singh, V. The effect of stimulant and depressant drugs on the latency of autokinetic illusion. *Acta Psychologica,* 1961, **18**, 354–359.

Skolnick, A. The role of eye movements in the autokinetic phenomenon. *Journal of Experimental Psychology,* 1940, **26**, 373–393.

Steinman, R. M., Cunitz, R. J., Timberlake, G. T., & Herman, M. Voluntary control of microsaccades during maintained monocular fixation. *Science,* 1967, **155**, 1577–1579.

Stoper, A. E. Vision during pursuit movement: The role of oculomotor information. Unpublished doctoral dissertation, Brandeis University, 1967.

Stutterheim, N. A. The primary position of the eyes. *British Journal of Ophthalmology,* 1933, **17**, 394–401.

Voth, A. C. Individual differences in the autokinetic phenomenon. *Journal of Experimental Psychology,* 1941, **29**, 306–322.

Voth, H. M. Choice of illness. *Archives of General Psychiatry,* 1962, **6**, 149–156.

Voth, H. M., & Mayman, M. A new dimension of personality organization. *Archives of General Psychiatry,* 1963, **8**, 366–380.

Wallach, H. The perception of motion. *Scientific American,* 1959, **200**(1), 56–60.

Walters, R. H., & Quinn, M. J. The effects of social and sensory deprivation on autokinetic judgments. *Journal of Personality,* 1960, **28**, 210–219.

Wapner, S., & Werner, H. *Perceptual development.* Worcester, Mass.: Clark University Press, 1957.

Weintraub, D. J., O'Connell, D. C., & McHale, T. J. Apparent vertical: Fundamental variables of sensory tonic theory reinvestigated. *Journal of Experimental Psychology,* 1964, **68**, 550–554.

Werner, H., & Wapner, S. Toward a general theory of perception. *Psychological Review,* 1952, **59**, 324–338.

Westheimer, G. Amphetamine, barbiturates, and accommodation convergence. *Archives of Ophthalmology,* 1963, **70**, 830–836.

Wetherill, G. B. *Sequential methods in statistics.* London: Methuen, 1966.

Whiteside, T. C. D., Graybiel, A., & Niven, J. I. Visual illusions of movement. *Brain: A Journal of Neurology,* 1965, **88**, 193–210.

Whitteridge, D. Central control of eye movements. In J. Field, W. H. Magoun, & V. E. Hall (Eds.), *Handbook of physiology.* (Section 1. Vol. 2.) Baltimore: Waverley Press, 1960.

Wieland, B. A., & Mefferd, R. B. Long term changes in properties of the autokinetic illusion. *Perceptual and Motor Skills,* 1966, **22**, 367–369.

Wishner, J., & Shipley, T. E. Direction of autokinetic movement as a test of the "sensory-tonic field" theory of perception. *Journal of Personality,* 1954, **23**, 99–106.

(Received July 6, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 6, 475

# THE HUDSON-DUNN CLUSTERING INDEX REVISITED

ROBERT L. HUDSON[1]

*Mississippi State University*

It is pointed out that the clustering index presented by Frankel and Cole in 1971 is identical to one previously published by Hudson and Dunn. A limitation of the Hudson-Dunn index is mentioned.

Frankel and Cole (1971) suggested a *Z*-score measure of category clustering based on the number of runs of similar items in a recall list. Although not mentioned by Frankel and Cole (1971), this index is the same as one previously presented by Hudson and Dunn (1969) and is similar to one presented by Dunn (1969). Hudson and Dunn (1969), however, based their index on the number of repetitions in a categorical list in keeping with the convention of using the repetition as the basic datum of clustering experiments. In a given list the number of repetitions and number of runs will bear a rather obvious relationship, and the *Z*-score calculated will be the same whether using number of repetitions or number of runs.

As both Hudson and Dunn (1969) and Frankel and Cole (1971) pointed out, the Hudson-Dunn index has some advantages over other indexes. It is not, however, a universal index. For example, the recall protocols *aabbccdd* and *aaaabbbbccccdddd* indicate perfect clustering of what has been recalled. The *Z*-score index for 1, however, is 3.24, while the *Z*-score index for 2 is 4.74. With more categories and more words recalled per category the difference is potentially much larger. Thus, the Hudson-Dunn index, along with most others, has limited usefulness when groups are being compared where there may be large differences in number of words recalled. This is particularly true when the number of categories represented may vary and when the amount of clustering approaches the maximum.

Other clustering indexes are available, but none of them would seem to be applicable to all research situations. Thus caution is advised in the selection and interpretation of clustering indexes in order to avoid misleading statements based on artifactual results.

## REFERENCES

DUNN, J. E. A compounded multiple runs distribution. *Journal of the American Statistical Association,* 1969, **64**, 1415–1423.

FRANKEL, F., & COLE, M. Measures of category clustering in free recall. *Psychological Bulletin,* 1971, **76**, 39–44.

HUDSON, R. L., & DUNN, J. E. A major modification of the Bousfield (1966) measure of category clustering. *Behavior Research Methods and Instrumentation,* 1969, **1**, 110–111.

(Received August 2, 1971)

---

[1] Requests for reprints should be sent to Robert L. Hudson, Department of Psychology, Mississippi State University, P.O. Drawer PF, State College, Mississippi 39762.

*Psychological Bulletin*
1972, Vol. 78, No. 6, 476–478

# COMMENT ON INTERPRETATION OF LATITUDE OF REJECTION AS AN "ARTIFACT"

CAROLYN W. SHERIF [1]

*Pennsylvania State University*

The use of latitude of rejection as an indicator of involvement is based on findings from two dissimilar procedures for attitude research. Interpreting this indicator as artifactual owing to its correlation with own position, Markley neglected one procedure, assumed equal intervals (untenable for both methods), ignored variability of the rejected latitude *across* positions, and constructed a threshold measure unrelated to the threshold concept in the research under review. Available research not reviewed supports the indicator of involvement for persons upholding positions other than the extremes, although as previously proposed the relative sizes of acceptable, objectionable, and noncommittal latitudes should serve as a more adequate indicator.

Several errors of interpretation and fact in Markley's (1971) article on latitude of rejection as an indicator of involvement call for correction.

Markley's article discussed the latitude of rejection as an indicator of personal involvement in an attitude object, confining his speculation to the method of ordered alternatives. The method of ordered alternatives is one technique for research on the frequency of judgments in *evaluative categories* as a function of the person's most acceptable position (own position) and relative involvement in the object of judgment (Sherif & Hovland, 1961; Sherif & Sherif, 1967; Sherif, Sherif, & Nebergall, 1965). Markley's critique assumed, for "didactic" purposes, that items used in this method had equal intervals for acceptance, rejection, and noncommitment, and attempted to demonstrate that the size of the latitude of rejection is "artifactually contaminated" owing to its correlation with extremity of own position.

The basic error is the author's persistence in assuming a model of scaling (equal intervals) that was never assumed in construction or use of the method of ordered alternatives, as he recognized (Markley, 1971, p. 359). This method was, in fact, an attempt to study attitude *without* certain arbitrary assumptions of conventional scaling models, particularly equal-appearing intervals, an assumption which, in Markley's words, "is not achieved in most applications and can never be perfectly validated [p. 359]."

Relationships between the latitude of rejection, own position, and ego-involvement (including those cited by Markley, 1971, p. 359) were based on data from the method of ordered alternatives and the own-categories procedure (Sherif et al., 1965). In *both* procedures, the size of the objectionable latitude is determined simply by the *relative* frequency of items rejected. In the own-categories procedure, only the extreme items are unambiguously ordered (low variability of judged position). The subject is free to use any number of categories and to distribute the items into them as he sees fit in order to group items whose position he sees as belonging together. Thus, Markley's didactic exercise of assuming equal item intervals for ordered alternatives (which did *not* make the assumption) would be equally erroneous for the own-categories procedure, which he regards as a "viable alternative [p. 359]."

The sources cited by Markley repeatedly emphasize relationships among the *relative* sizes of acceptable, objectionable, and noncommittal latitudes, rather than the absolute frequency within any single latitude. Nevertheless, the latitude of rejection was used as an indicator of involvement on the basis of findings from *both* methods, which were in agreement. Markley cited only one of the conclusions supporting latitude of rejection as an indicator of involvement, that is, the correlation between extremity of position and latitude of rejection. The inference that this correlation was related to greater involvement of extreme subjects was, in turn, based on (*a*) the well-known U curve relating extremity and intensity ratings (e.g., Cantril, 1946) and (*b*) the large latitudes of rejection typically obtained from extreme subjects actively committed to upholding their stands (Sherif et al., 1965, p. 29).

---

[1] Requests for reprints should be sent to Carolyn W. Sherif, Pennsylvania State University, 417 Psychology Building, University Park, Pennsylvania 16802.

The second finding supporting latitude of rejection as an indicator of involvement was its *variability* across attitude positions from extremes through moderate own positions (Sherif et al., 1965, p. 156). Thus, another error in Markley's comments lies in constructing a "didactic" model of three response sets for different extremities in own position (Table 1, p. 358) without considering variability among persons upholding the *same* position. In a study of reactions to communications on a national election, there were sufficient numbers of subjects adopting *each* of nine positions (extreme through moderate) who had latitudes of rejection composed of five or more positions (as did only the *extreme* response set in Markley's Table 1) to compare their ratings of communications with subjects adopting *each* of nine positions who rejected four or fewer items (Sherif et al., 1965, pp. 156 ff.). Designating subjects *across* the extreme and moderate positions as high in involvement if they rejected five or more positions, the research found significant differences between their judgments of communications and those by less involved subjects (latitude of rejection of four or fewer items).

The fact that subjects adopting moderate positions do, in substantial numbers, exhibit latitudes of rejection as large as the typical extreme subjects is a blow to the main thrust of Markley's critique based on the unwarranted assumption of equal item intervals. His key proposition is that for the person adopting a position at one extreme "the latitude of rejection . . . is spuriously lengthened due to the increased distance to the most rejected item [p. 358]." The proposition implies that persons taking extreme positions reject more items simply because there is more "room" to do so between their own position and the most rejected item. Markley seems to forget that there is no logical or procedural requirement in the ordered alternatives method that the extreme subject do so. Instructions require *only* that he indicate his most acceptable and objectionable positions, leaving him free to reject or accept others if he wishes or *not* to respond to other items (remain noncommittal).

Many highly involved persons taking extreme positions do respond as in Markley's Response Set 1 (Markley, 1971, Table 1), rejecting *all* opposing positions *and* the position E (see Markley's Table 1), which is a noncommittal statement. If they do so because of an "artifact," such highly involved persons should respond differently when using the own-categories procedure. Yet with that procedure, such highly involved subjects typically employ very few categories and construct an extremely broad objectionable category, lumping all of the items opposing their own stand together. The proposition that the wide latitude of rejection thus obtained is "artifactual" is all the more difficult to sustain, as an equal-interval assumption is equally untenable, and the items are arranged in random order so that the subject may encounter the most objectionable item in any serial position.

It is not entirely clear why Markley coupled his proposition about distance to the most objectionable position with the possibility of a ceiling effect for the subject adopting the most extreme position. It is difficult to envision even a Republican zealot endorsing a more extreme position than Position A in Markley's Table 1 (p. 358), namely: "The election of the Republican presidential and vice presidential candidates in November is absolutely essential from all angles in the country's interests." In fact, even this position was so extreme that comparatively few of the subjects who were avowed and active supporters of the Republican candidates adopted it as most acceptable. The bulk of such committed Republican supporters adopted the adjacent Position B as most acceptable, with *more* than 50% of them failing to endorse the extreme Position A, and nearly 20% *rejecting* this adjacent, more extreme position (Sherif et al., 1965, p. 37). Therefore, the addition of a still more extreme position in the ordered alternatives might well have *increased* the latitude of rejection by providing an alternative "too extreme" even for the extremists.

Markley's interpretation of "threshold" of rejection, in terms of number of positions between the most acceptable position and the first rejected position, is not only "novel." It has nothing to do with the use of the threshold concept in the references cited, which refers to the probability that a stimulus will be placed in a given category rather than an adjacent or other category. This is the usual definition of the difference threshold in psychophysical research. Far from being "essentially an interval measure [p. 359]," the difference threshold is a means of defining *limits* of psychological categories in terms of stimulus values. True, the stimulus values in most psychophysical research are arrayed in equal intervals, but application of the threshold concept to problems of *evaluative* categories (acceptance, rejection, noncommitment) need not make this assumption in view of the fact that evaluations are *necessarily* human judgments and need not follow the regularities of physical dimensions.

If Markley's basic exercise, assuming equal intervals, and his conclusion of an "artifactual contamination" of latitudes of rejection for *extreme* subjects were valid, by the same token moderate subjects should be subject to the converse artifact, preventing them from rejecting as many positions because their own position is close to both extremes. As noted above, substantial numbers of moderate subjects do reject five or more out of nine positions. In work published before Markley's (1971) article was received for publication, there is further evidence that a moderate or noncommittal subject *can* be highly involved and reject items as frequently as extreme subjects (Sherif & Sherif, 1969). For example, subjects noncommittal on the choice of available candidates in an election (adopting Position E in Markley's Table 1, p. 358), but active and highly involved in the outcome, display latitudes of rejection comparable in size to those upholding extreme positions (adopting Position A in Table 1, p. 358). Politically inert subjects upholding the same position (Position E) typically evaluate the alternatives as in Markley's Response Set 3, rejecting fewer positions. Similarly, those siding *moderately* with one candidate or another, whether Republican or Democratic, reject significantly more items if they are politically active than if politically inactive (Sherif & Sherif, 1969, p. 365). These, and several other more recent investigations, permit the conclusion that subjects upholding positions other than the extremes can be highly involved in them and that when they are, they exhibit latitudes of rejection comparable to those of the typical subject upholding an extreme position.

Hopefully, these comments may indicate some dangers in a selective review based on criteria (psychometric scaling and equal-appearing intervals) that the research under review explicitly rejects as stifling progress in the important but muddied problem area of attitude and attitude change.

## REFERENCES

Cantril, H. The intensity of an attitude. *Journal of Abnormal and Social Psychology*, 1946, **41**, 129–135.

Markley, O. W. Latitude of rejection: An artifact of own position. *Psychological Bulletin*, 1971, **75**, 357–359.

Sherif, M., & Hovland, C. I. *Social judgment: Assimilation and contrast effects in communication and attitude change.* New Haven: Yale University Press, 1961.

Sherif, M., & Sherif, C. W. Attitude as the individual's own categories: The social judgment-involvement approach to attitude and attitude change. In C. W. Sherif & M. Sherif (Eds.), *Attitude, ego-involvement and change.* New York: Wiley, 1967.

Sherif, M., & Sherif, C. W. *Social psychology.* New York: Harper & Row, 1969.

Sherif, C. W., Sherif, M., & Nebergall, R. *Attitude and attitude change: The social judgment-involvement approach.* Philadelphia: Saunders, 1965.

(Received June 1, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 6, 479

# REJOINDER TO SHERIF

O. W. MARKLEY[1]

*Educational Policy Research Center, Stanford Research Institute, Menlo Park, California*

The scaling artifact intrinsic to the method of ordered alternatives, as presented by Markley in 1971, is defended from the comments by Sherif. The essential conclusion of the 1971 articles remains unchanged.

Sherif's (1972) detailed attack on my critique (Markley, 1971) of the method of ordered alternatives is interesting but, for the most part, specious. Her comments regarding "equal intervals," however, reveal a central misunderstanding of the analysis.

In constructing the set of "didactic responses" on which the critique was based, the central assumption was not that each item of the *scale* had an equal interval on a dimension of extremity, but rather that each of three sets of *responses* (which differed in extremity) had equal intervals for acceptance, noncommitment, and rejection of items. Thus the assumptions underlying the method of ordered alternatives were not violated.

I did not deny that extremity and involvement are usually related, nor that differences in involvement are reflected by differences in the latitude of rejection. The problem is simply that this reflection is contaminated by a scaling artifact intrinsic to the scoring of the method of ordered alternatives. Thus this method is a dubious measure of involvement across respondents of differing extremity—a conclusion that still seems valid.

## REFERENCES

Markley, O. W. Latitude of rejection. An artifact of own position. *Psychological Bulletin,* 1971, **75,** 357–359.

Sherif, C. W. Comment on interpretation of latitude of rejection as an "artifact." *Psychological Bulletin,* 1972, **78,** 476–478.

(Received June 1, 1971)

---

[1] Requests for reprints should be sent to O. W. Markley, Stanford Research Institute, Menlo Park, California 94025.

task. While this has advantages over the yoked situation, the assumption that an animal performing a task already learned is not undergoing new learning need not be true. For example, Mackintosh (1965) has found that rats given a moderate amount of training in a T-maze position problem respond primarily on the basis of "place" or environmental cues, while rats given further training on the problem come to respond on the basis of "response," or internal kinesthetic cues. Thus, while performing at a high level of accuracy, the rats in the Mackintosh experiment learned something new. It might be argued that the rats that have no contingency change undergo less new learning than rats given a new task. (Indeed, most psychologists would probably accept this.) Unfortunately, such an argument is difficult to make empirically in the absence of a common metric on which the amounts of different kinds of learning may be compared.

The above problems, or similar ones, seem to be present in all of the learning models that Gaito and Bonnet discuss. It seems that the elucidation of the macromolecular correlates of learning and memory will require more careful attention to behavioral techniques than has generally been applied. The current approach requires an animal that is not learning or is learning considerably less than the experimental subject but that undergoes the nonlearning aspects of the training situation. While the "overtrained control" paradigm (Bowman & Strobel, 1969) may approach such a situation, it seems impossible on logical grounds to equate all sensory, motor, motivational, etc., aspects of situations in which the animals are learning different things. It is far from clear that an ideal behavioral paradigm exists for the detection of molecular changes resulting from (in) learning and memory, but it is clear that the traditional procedures used in the studies reviewed by Gaito and Bonnet, and in particular the yoked control, are conceptually unsatisfactory.

## REFERENCES

ADAIR, L. B., WILSON, J. E., & GLASSMAN, E. Brain function and macromolecules, IV. Uridine incorporation into polysomes of mouse brain during different behavioral experiences. *Proceedings of the National Academy of Sciences,* 1968, **61**, 917–922.

AZMITIA, E. C., & MCEWEN, B. S. Corticosterone regulation of tryptophan hydroxylase in midbrain of the rat. *Science,* 1969, **166**, 1274–1276.

BOWMAN, R. E., & STROBEL, D. A. Brain RNA metabolism in rat during learning. *Journal of Comparative and Physiological Psychology,* 1969, **67**, 448–456.

CHURCH, R. M. Systematic effect of random error in the yoked control design. *Psychological Bulletin,* 1964, **62**, 122–131.

GAITO, J., & BONNET, K. Quantitative versus qualitative RNA and protein changes in the brain during behavior. *Psychological Bulletin,* 1971, **75**, 109–127.

HYDÉN, H., EGYHÁZI, E., JOHN, E. R., & BARTLETT, F. RNA base ratio changes in planaria during conditioning. *Journal of Neurochemistry,* 1969, **16**, 815–821.

JAKOUBEK, B., SEMIGINOVSKÝ, B., KRAUS, M., & ERDOSSOVÁ, R. The alterations of protein metabolism of the brain cortex induced by anticipation stress and ACTH. *Life Sciences,* 1970, **9**, 1169–1179.

JOHN, E. R. *Mechanisms of memory.* New York: Academic Press, 1967.

MACHLUS, B., & GAITO, J. The use of successive competition hybridization to detect RNA species in a shock avoidance task. *Nature,* 1969, **222**, 573–574.

MACKINTOSH, N. J. Overtraining, transfer to proprioceptive control and position reversal. *Quarterly Journal of Experimental Psychology,* 1965, **17**, 26–36.

MAIER, S. F., SELIGMAN, M. E. P., & SOLOMON, R. L. Pavlovian fear conditioning and learned helplessness: Effects on escape and avoidance behavior of (a) CS-US contingency and (b) the independence of the US and voluntary responding. In B. A. Campbell & R. Church (Eds.), *Punishment.* New York: Appleton-Century-Crofts, 1969.

RESCORLA, R. A. Pavlovian conditioning and its proper control procedrues. *Psychological Review,* 1967, **74**, 71–80.

SELIGMAN, M. E. P. Control group and conditioning: A comment on operationism. *Psychological Review,* 1969, **76**, 484–491.

SINGH, D. Preference for bar pressing to obtain reward over freeloading in rats and children. *Journal of Comparative and Physiological Psychology,* 1970, **73**, 320–327.

WAGNER, A. R. Elementary associations. In H. H. Kendler & J. T. Spence (Eds.), *Essays in neobehaviorism: A memorial to Kenneth Spence.* New York: Appleton-Century-Crofts, 1971, in press.

WEISS, J. M. Effects of coping responses on stress. *Journal of Comparative and Physiological Psychology,* 1968, **65**, 251–260.

WEISS, J. M., STONE, E. A., & HARRELL, N. Coping behavior and brain norepinephrine level in rats. *Journal of Comparative and Physiological Psychology,* 1970, **72**, 153–160.

(Received June 17, 1971)

*Psychological Bulletin*
1972, Vol. 78, No. 6, 483-484

# CONTROLS IN MOLECULAR PSYCHOBIOLOGICAL RESEARCH

JOHN GAITO[1]
*York University*

KENNETH BONNET
*Stanford Research Institute*

The problem of adequate controls in molecular psychobiological research is acknowledged. Three methodological possibilities that attempt to handle this aspect are discussed.

We agree with Greenough and Maier (1972) that the provision of controls that are conceptually adequate presents a problem in molecular psychobiological research. This is an aspect with which we have struggled since the inception of our research program a number of years ago. In some cases, we (and others) allowed for the possibility that the "nonlearning" animal might be learning something, and assumed that the learning in this animal was minimal. However, to more adequately handle this problem, we have considered three behavioral methodologies in which the neurochemical contributions underlying particular behavioral experiences would be evaluated.

Since we are only beginning to determine the type of biochemical systems in the brain that respond to behavioral treatment, one approach has been to maximize behavioral differences between animals used for neurochemical analysis so as to locate potential differences. If chemical differences were not evident, further behavioral variations would be introduced. However, if neurochemical differences were found, the investigator would then systematically determine the contributions for each of a number of specific behavioral components in the original experimental treatment. Ultimately, one would hope to reach a point at which a particular chemical event in the brain might be attributed to the learning process, or to what ever behavioral event was of primary concern.

This is the behavioral strategy originally taken in our deoxyribonucleic acid-ribonucleic acid (DNA-RNA) hybridization experiments. In the first experiment (Machlus & Gaito, 1968a), the control animals remained in the home cage, and the experimentals received shock avoidance training. In further experiments controls were involved in a motor task only, or were yoked to the experimental animal during training (Machlus & Gaito, 1968b, 1969). Hybridization studies indicated chemical events unique to the avoidance-trained animal in each case. Other individuals have also obtained a specific chemical result, then proceeded to determine the behavioral specificity of that result (e.g., Glassman and his research unit at the University of North Carolina).

Another possibility is to use the "degree of training" method. Animals would be sacrificed at various stages or times during the acquisition phase of learning, at various times after acquisition is complete, and during later performance of the task learned. Changes in neurochemical systems over time would be of interest and would be related to various phases in the development and stabilization of the learning process. This approach has been used previously by a number of investigators (e.g., Hyden & Lange, 1965); recently we have implemented this methodology also.

A third approach uses the surgically split-brain animal in a monocularly presented discrimination task. The "nonlearning" hemisphere provides a control for motoric, somesthetic, and affective aspects of the training experience. A series of recent studies from this laboratory (Bonnet, 1971) may suggest that monocular visual discrimination training effects qualitative alterations in brain genome transcription which are restricted to the "trained" hemisphere.

A major point that Greenough and Maier (and others) overlook is that many researchers in this area are not attempting to "prove[2] either the presence or absence of molecular events peculiar to learning [Greenough & Maier, p. 480]" by *one or a few experiments;* a number of research teams have *systematic programs using different behavioral and molecular biological techniques with the expectation that ultimately the results will point rather conclusively to particular neurochemical events or systems which are responsive to behavioral events.* Few programs in this area of research have been developed to the point of testing of major premises or hypotheses at this time, though nearly all have some guiding theoretical orientation.

---

[1] Requests for reprints should be sent to John Gaito, Department of Psychology, York University, 4700 Keele Street, Downsview 463, Ontario, Canada.

[2] The use of the word "prove" by Greenough and Maier is inappropriate because in empirical science we never prove anything; we merely make one

One specific implication of the Greenough and Maier critique lies in the dichotomizing of "learning" and "nonlearning." While we agree that the control animal in nearly all treatments usually given is probably learning something, the "learning" resulting from greater novel contingency in the experimental treatment situation is implied to be a unitary phenomenon. Yet, most learning as encountered in the work of experimental psychologists probably involves the participation of numerous nervous structures, cell types, and neural events. These in turn are probably subserved by a variety of biochemical mechanisms, each responsive to a particular phase or component of the "learning" event.

Finally, we agree with Greenough and Maier that behavioral techniques require more careful attention in molecular psychobiological research. The aim and types of concerns expressed are pertinent to the area of biochemical substrates of behavior, and point to the need for furthering the development of design technology in this burgeoning area.

---

alternative appear more and more certain using probabilistic methodology.

## REFERENCES

Bonnet, K. A. A biochemical response to unilateral visual stimulation in the split-brain preparation. Unpublished doctoral dissertation, York University, Toronto, 1971.

Greenough, W. T., & Maier, S. F. Molecular changes during learning: Behavioral strategy—a comment on Gaito and Bonnet. *Psychological Bulletin,* 1972, **78**, 480–482.

Hyden, H., & Lange, P. W. A differentiation in RNA response in neurons early and late during learning. *Proceedings of the National Academy of Sciences,* 1965, **53**, 946–952.

Machlus, B., & Gaito, J. Detection of RNA species unique to a behavioral task. *Psychonomic Science,* 1968, **10**, 253–254. (a)

Machlus, B., & Gaito, J. Unique RNA species developed during a shock avoidance task. *Psychonomic Science,* 1968, **12**, 111–112. (b)

Machlus, B., & Gaito, J. The use of successive competition hybridization to detect RNA species in a shock avoidance task. *Nature,* 1969, **222**, 573–574.

(Received October 15, 1971)

ABILITATION-INDIA